



Les théorèmes ergodiques en simulation

Mohamed Ben Alaya

► To cite this version:

Mohamed Ben Alaya. Les théorèmes ergodiques en simulation. Sciences de la Terre. Ecole Nationale des Ponts et Chaussées, 1992. Français. NNT: . tel-00523186

HAL Id: tel-00523186

<https://pastel.archives-ouvertes.fr/tel-00523186>

Submitted on 4 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

79961

NS 16737

(2)

Ecole Nationale des Ponts et Chaussées

Thèse de doctorat

spécialité

Mathématiques appliquées.

sujet de thèse :

LES THEOREMES ERGODIQUES EN SIMULATION.

Présentée par Mohamed Ben Alaya

et soutenue le 11 décembre 1992

devant le jury composé de :

Jean François Méla (Président)

Gilles Pagès (Rapporteur)

Denis Talay (Rapporteur)

Nicolas Bouleau

Dominique Lepingle

Etienne Pardoux



EN.P.C.



DOC09151

02

Mes pensées vont tout d'abord à ma famille qui m'a toujours encouragé à entreprendre des études valorisantes. Je voudrais leur exprimer mes remerciements les plus sincères. Ils m'ont donné accès à une éducation exceptionnelle, tout en m'offrant des occasions uniques, quoique difficiles à saisir à l'époque, de me réaliser par moi-même.

A Nicolas Bouleau, je voudrais faire part de ma reconnaissance pour m'avoir ouvert, en m'accueillant au CERMA, des horizons de recherche dans l'univers des mathématiques appliquées. Il a eu la lourde tâche de diriger ma thèse. Je tiens à lui exprimer ma profonde gratitude et à le remercier très chaleureusement.

Gilles Pagès, a été sans le savoir d'une aide précieuse. C'est tout d'abord avec lui et pendant mon mémoire de DEA que j'ai commencé à découvrir le domaine des probabilités numériques. Il a par ailleurs su m'encourager dans mon travail, et ce, de nombreuses fois. Je lui dois de nombreuses discussions passionnantes et constructives. Je suis donc très heureux et reconnaissant qu'il ait bien voulu accepter la tâche ingrate de rapporteur. Je l'en remercie vivement.

Denis Talay a toujours exprimé beaucoup d'intérêt pour les probabilités numériques. Je lui dois de nombreuses invitations à l'INRIA où la collaboration avec lui a été très agréable. Il a sans cesse témoigné de l'intérêt pour mon travail. Je suis enchanté qu'il ait accepté d'être rapporteur et je le remercie chaleureusement.

Jean-François Méla me fait l'honneur de présider ce jury. J'apprécie particulièrement sa compétence et la sympathie qu'il m'a toujours exprimé. Qu'il soit assuré de ma reconnaissance.

Quiconque connaît Etienne Pardoux, la diversité et l'originalité de ses travaux est honoré de le voir parmi les membres de son jury. Je suis très flatté qu'il soit présent et je le remercie chaleureusement.

Dominique Lepingle a accepté de participer à ce jury. Je le remercie très vivement de l'honneur qui m'est ainsi fait.

Bernard Lapeyre et Damien Lambertson m'ont toujours prêté une oreille attentive. Je les ai de nombreuses fois importuné par des questions et ils m'ont de nombreuses fois patiemment répondu, que je leur présente mes excuses et mes remerciements les plus sincères.

J'adresse mes remerciements chaleureux pour Monia Bellalouna, Stéphane Jaffard, Serge Cohen et Zhang Xiao-lan d'avoir toujours su entretenir autour de moi une atmosphère amicale qui a eu sa part dans la réalisation de ce travail.

Durant ces années, j'ai pu jouir de l'excellente ambiance qui règne au CERMA, je remercie l'ensemble, les autres chercheurs et les thésards avec qui j'ai partagé les mêmes doutes et les mêmes questions. Véronique Serre est plus particulièrement responsable de ce qu'on ne voit pas dans ces pages. Elle contribue à faire de ce centre un lieu de travail convivial, je la remercie très chaleureusement.

Enfin je remercie des êtres chers, qui se reconnaîtront dans ces lignes, et savent l'importance qu'a eu et a encore leur amitié ou leur amour.

Table des Matières

Introduction	4
Partie I La METHODE DU DECALAGE	7
1 Rappel sur la méthode de Monte Carlo.	8
1.1 Simulation de variables aléatoires.	8
1.2 La loi forte des grands nombres.	9
1.3 Vitesse de convergence de la méthode de Monte Carlo.	10
2 Présentation de la méthode du décalage	12
2.1 Cadre mathématique de la méthode.	12
2.2 Principe du dispositif informatique.	14
2.2.1 Quelques résultats numériques sur l'efficacité du shift.	17
2.3 Les résultats existants sur la vitesse de convergence.	20
3 Vitesse de convergence presque sûre.	22
3.1 Théorème de Gál et Koksma et vitesse de convergence dans le théorème de Birkhoff.	22
3.2 Sur la vitesse de Convergence du shift.	29
3.3 Vitesse de convergence des fonctions \mathcal{F}_T mesurables.	34
3.4 Illustration informatique.	37
3.5 Conclusion.	40
4 Convergence en loi de l'erreur normalisée.	42
4.1 Théorème de la limite centrale pour les fonctions dépendant d'un temps d'arrêt.	42
4.2 Théorème de la limite centrale pour les fonctions pratiquement cylindriques.	47
4.3 Théorème de la limite centrale pour les fonctions lipschitziennes.	49
4.3.1 L'opérateur de Perron-Frobenius et ses perturbations.	49
4.3.2 Théorème de la limite centrale.	55
4.3.3 La vitesse dans le théorème de la limite centrale.	58
4.4 Illustration informatique.	60
4.5 Conclusion.	63

5	Sur la loi du logarithme itéré	64
5.1	Sur la loi du logarithme itéré pour l'opérateur de décalage.	64
5.1.1	Résultats théoriques existants pour les fonctions pratiquement cylindriques.	65
5.1.2	Formulation dans le cas des fonctions \mathcal{F}_T -mesurables.	67
5.2	Sur la loi du logarithme itéré pour l'opérateur de décalage à droite	69
5.2.1	Classe de Gordin et finitude de σ^2	73
5.2.2	Vitesse de convergence des fonctions \mathcal{F}_T mesurables pour le shift à droite.	75
5.2.3	Vitesse de convergence des fonctions pratiquement cylindriques pour le shift à droite.	77
5.3	Conclusion	79
6	Résolution des équations elliptiques par la méthode du shift	80
6.1	Représentation probabiliste des solutions d'une équation aux dérivées partielles	80
6.2	Résultats numériques	82
6.3	Interprétation des résultats obtenus	84
6.4	Application de la méthode du shift à la résolution des équations elliptiques.	88
6.5	Illustration informatique.	90
6.6	Emploi de la méthode du shift sur d'autres représentations du mouvement brownien.	92
Partie II	Approximation des mesures invariantes	94
7	Calcul d'une intégrale par rapport à une mesure invariante.	95
7.1	Présentation de la méthode ergodique.	95
7.1.1	Rappel sur la récurrence des chaînes de Markov à espace d'états dénombrable.	96
7.1.2	Rappel sur la récurrence des chaînes de Markov à espace d'états non dénombrable.	97
7.1.3	Utilisation de la méthode par les physiciens.	98
7.2	Exposant de Lyapunov.	99
7.2.1	Approximation des exposants de Lyapunov.	101
7.2.2	Résultats numériques	103
7.3	Conclusion.	108

Introduction

Pour situer cette thèse dans son contexte, on peut la voir comme un pont jeté entre théorie et applications véritables avec pour souci permanent de justifier les algorithmes présentés et leurs propriétés par des démonstrations mathématiques rigoureuses. En général, il faut reconnaître qu'il y a un décalage surprenant entre l'énorme littérature théorique consacrée aux probabilités et les applications qui utilisent effectivement cette littérature. Le domaine des **probabilités numériques** essaie d'articuler les deux ; il peut être vu comme un maillon indispensable entre la théorie et les applications. Cette thèse en fait partie. On propose dans ce travail des algorithmes de simulations accompagnés à la fois de résultats portant sur leurs performances observées ainsi que des démonstrations mathématiques expliquant ces performances. Le corpus mathématique utilisé fait intervenir des concepts probabilistes ainsi que des théorèmes et des techniques relevant de la théorie ergodique. Tout ceci avec pour préoccupation d'améliorer ou de construire des méthodes numériques réalistes et efficaces.

La thèse se compose de deux parties indépendantes. La première est consacrée à l'étude de la méthode du décalage, dite aussi méthode du shift, pour le calcul d'espérances mathématiques en dimension grande ou infinie. La deuxième partie s'attache au problème de l'approximation des mesures invariantes pour les chaînes de Markov.

La méthode de Monte Carlo pour le calcul d'espérances mathématiques, c'est à dire l'implémentation informatique de la loi forte des grands nombres, est connue et utilisée depuis longtemps dans de très nombreuses applications des sciences de l'ingénieur où le hasard intervient, aussi bien que dans des phénomènes physiques et économiques très variés. Dans le premier chapitre nous commençons par quelques rappels. Il s'agit ici d'une présentation rapide et élémentaire qui s'appuie sur la simulation des variables aléatoires et son application aux calculs d'espérances mathématiques. Au passage, les résultats classiques sur la vitesse de convergence de la méthode de Monte Carlo sont rappelés.

Cependant, l'usage du théorème ergodique ponctuel de Birkhoff pour le calcul d'espérances mathématiques, à cause des particularités de son implémentation informatique se trouve être, dans de nombreuses situations, plus efficace que la méthode de Monte Carlo "classique", quoique cette méthode ne soit pas encore répandue. Dans le chapitre 2, on rappelle d'abord quelques résultats de la théorie ergodique puis on présente la méthode du décalage (ou du shift), pour le calcul d'espérances en dimension grande ou infinie (voir [1], [2] et [12]). Pour l'essentiel, la méthode du décalage est la mise en œuvre informatique du théorème ergodique ponctuel pour l'opérateur du décalage (à gauche ou, à défaut, à droite). L'espérance mathématique est simulée sur des trajectoires dépendantes ; les performances de ce procédé proviennent d'une "boîte de stockage" qui évite de recommencer inutilement certains calculs partiels en passant d'une trajectoire à une autre et qui permet, donc, une vitesse de convergence informatique plus grande que la méthode de Monte Carlo.

L'utilisation de cette méthode nécessite comme Monte Carlo des estimations de la vitesse de convergence, telles que la loi du logarithme itéré et le théorème de la limite centrale, dont le théorème ergodique ponctuel général ne dispose pas (cf. Krengel [32]). Dans le but de donner des critères d'arrêts effectifs pour ce type d'algorithmes, je me suis intéressé à l'obtention de vitesses de convergence pour une certaine classe de fonctions que l'on rencontre fréquemment en pratique : la classe des fonctionnelles F de carré intégrable et mesurables par rapport à la tribu \mathcal{F}_T des événements antérieurs à un temps d'arrêt T .

Le chapitre 3 commence par quelques rappels sur l'énoncé d'un théorème de Gál et Koksma [18], sa démonstration dans un cadre probabiliste et sa formulation dans le cadre de l'opérateur du shift. La suite du chapitre est consacrée à une "loi du logarithme itéré faible" qui donne une vitesse de convergence dans le théorème ergodique de Birkhoff dans le cas de transformations fortement mélangeantes. Ce résultat, appliqué à la transformation du shift, permet d'obtenir une vitesse de convergence du même ordre en pratique que la méthode de Monte Carlo pour une certaines classes raisonnablement générales de fonctionnelles F couramment employées en pratique.

Afin d'étudier le comportement asymptotique en loi de l'erreur normalisée de cette méthode, on établit au chapitre 4 des théorèmes de limites centrales. Dans une première partie de ce chapitre et en utilisant des techniques similaire à celles de Billingsley [9], on démontre le théorème de la limite centrale pour les classes de fonctions introduites au chapitre précédent. Dans une deuxième partie, on s'appuie cette fois sur des techniques d'analyse fonctionnelle pour obtenir le même résultat pour les fonctions lipschitziennes. Cette dernière classe de fonctions est moins utile dans le cadre de nos applications, cependant, la démonstration nous paraît intéressante et permet en outre d'obtenir une vitesse de convergence dans le théorème de la limite centrale de type Berry-Essen.

Au chapitre 5 nous présentons d'abord quelques résultats sur la loi du logarithme itéré de Walter Phillip et William Stout [42] puis leur formulation dans le cadre de l'opérateur du shift ; ceci nous permet d'obtenir la loi du logarithme itéré pour une classe de fonctions strictement contenue dans celle vérifiant la loi "faible" au chapitre 3, la restriction s'exprime en conditions sur les moments. L'étude de la vitesse de convergence dans le théorème de Birkhoff pour l'opérateur de décalage à droite semble mieux adaptée aux résultats théoriques existants (voir N. Bouleau [13]) ; ceci fait l'objet de la dernière partie de ce chapitre. Dans un premier temps sont rappelés les résultats obtenus par N. Bouleau sur la vitesse de convergence de la méthode du décalage à droite. En particulier, il est montré que, pour une certaine classe de fonctions, dite classe de Gordin, le théorème de la limite centrale et la loi du logarithme itéré sont vérifiés. Pour notre part, nous montrons que les classes de fonctionnelles introduites au chapitre 3 sont dans la classe de Gordin ; ceci améliore le résultat de N. Bouleau sur les fonctions dépendant d'un temps d'arrêt.

La conclusion essentielle de ces trois derniers chapitres est que la méthode du décalage a une vitesse de convergence asymptotique (mathématique) du même ordre de grandeur, $O(\frac{1}{\sqrt{n}})$, que la loi des grands nombres, pour une large classe de fonctionnelles parmi les plus couramment rencontrées. Cette vitesse de convergence est contrôlée par un paramètre σ^2 , représentant la somme de la variance et des corrélations entre les trajectoires décalées ; le rôle de σ^2 est similaire à celui joué par la variance dans la méthode de Monte Carlo.

Le chapitre 6 est une application de cette méthode à la résolution d'équations elliptiques. On observe notamment que la vitesse de convergence de l'algorithme est contrôlée par le

paramètre σ^2 qui lui, dépend de la représentation du mouvement brownien et de la transformation ergodique choisie. En effet, pour une discrétisation du brownien par le schéma d'Euler "classique" et pour diverses transformations, on observe des vitesses de convergence complètement différentes, une application naïve de la méthode pouvant aboutir à des problèmes ; une estimation du terme σ^2 , dans chaque cas, permet d'expliquer ce phénomène. A la fin de ce chapitre nous proposons d'autres représentations du mouvement brownien dont nous pensons que l'usage avec la méthode du shift pourraient se révéler efficace que le schéma d'Euler.

La deuxième partie de cette thèse, constituée du chapitre 8, s'attache au problème de l'approximation des mesures invariantes. Il s'agit d'une présentation élémentaire des chaînes de Markov ergodiques suivie d'un panorama sur l'utilisation de cette théorie par les physiciens ; on trouve en outre leur point de vue sur la vitesse de convergence de ce type d'algorithme. Une application aux calculs des exposants de Lyapunov est donnée à la fin de ce chapitre. Ceci se situe dans le droit fil des préoccupations de D. Talay[52][20].

Pendant ma thèse, je me suis servi des simulations numériques comme outils d'expérimentation permettant d'illustrer les résultats obtenus et de tester l'efficacité informatique des méthodes proposées, autrement dit de confirmer, d'infirmer, ou de découvrir de nouvelles propriétés mathématiques. On trouvera, donc dans chaque chapitre, des résultats numériques accompagnés de commentaires.

Partie I

La METHODE DU DECALAGE

Chapitre 1

Rappel sur la méthode de Monte Carlo.

1.1 Simulation de variables aléatoires.

En théorie, toute variable aléatoire X peut être mise sous la forme $X = \varphi(U)$ où φ est dx -intégrable sur $[0, 1]$ et U est une variable aléatoire de loi uniforme sur $[0, 1]$: on prend

$$\varphi(U) = \inf\{x / F(x) \geq u\}$$

où $F(x) = \mathbb{P}(X \leq x)$ est la fonction de répartition de X . A priori, simuler une suite de variables aléatoires donnée revient donc à simuler une suite de nombres au hasard, plus précisément une réalisation d'une suite de variables aléatoires $U_1, U_2, \dots, U_n, \dots$ indépendantes de loi uniforme sur $[0, 1]$. Malheureusement, ce problème est mathématiquement difficile et il n'est pas effectif puisqu'il exclut toutes les suites obtenues par un algorithme donné (cf. Dellacherie et Martin Lof [11]). Cependant on peut construire des nombres, dits pseudo-aléatoires, vérifiant plusieurs tests statistiques. Ces derniers ont fait l'objet de nombreux travaux (cf. par exemple [4]). La précision des ordinateurs étant finie, dans la pratique on n'a accès qu'à des suites finies de nombres pseudo-aléatoires que l'on assimile à une réalisation $U_1(\omega), U_2(\omega), \dots, U_N(\omega), \dots$ d'une suite de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Un procédé très employé pour construire ces suites est de considérer des suites x_n à valeurs entières entre 0 et m et de prendre $U_n = \frac{x_n}{m}$. On peut définir les x_n par récurrence à partir d'une formule de type

$$x_{n+1} = ax_n \bmod m \quad \text{ou} \quad x_{n+1} = ax_n + b \bmod m$$

où a et b sont des entiers et m est un entier assez grand. En prenant a et m premiers entre eux on obtient une suite de période r , avec r le plus petit entier tel que $a^r = 1$.

Donc, à première vue la simulation de variables aléatoires de loi $\mathcal{L}(X)$ ne semble pas poser de problème particulier ! Cependant, sauf dans des cas très simples (loi exponentielle, somme d'uniformes, lois à supports finis) φ n'est pas numériquement accessible. En outre la connaissance explicite de φ étant clairement équivalente à celle de la loi de X (via F) il est bien évident qu'une telle situation ne se rencontre que rarement dans les problèmes réels. Par contre cette technique peut être utilisée pour simuler des variables intermédiaires.

En effet, le plus souvent, une variable aléatoire X apparaît naturellement sous la forme :

$$X = f(U_1, \dots, U_d, \dots)$$

où U_1, \dots, U_d, \dots est une suite, finie ou non, de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$ et f une fonction intégrable de \mathbb{R}^d dans \mathbb{R} ou de $\mathbb{R}^{(\mathbb{N})} = \cup_{n \geq 1} \mathbb{R}^n$ dans \mathbb{R} . La fonction f est généralement définie à partir d'un algorithme : elle traduit mathématiquement à la fois le système (physique, économique, etc) qui engendre X et les méthodes de simulation utilisées pour engendrer d'éventuelles variables intermédiaires (dans le cas de l'emploi de la méthode de rejet par exemple). Nous renvoyons à N. Bouleau [11] pour la simulation de variables aléatoires.

La représentation de X par la fonction f sur $\mathbb{R}^{(\mathbb{N})}$ nous permettra grâce à la loi forte des grands nombres de déterminer l'espérance de X , $\mathbb{E}(X)$.

1.2 La loi forte des grands nombres.

Le calcul d'espérances mathématiques est l'un des pôles essentiels des probabilités numériques. D'un point de vue théorique le calcul d'espérance d'une variable aléatoire réelle X est un simple calcul d'intégrale

$$\mathbb{E}X = \int_{\mathbb{R}} x \mu(dx)$$

où μ (la loi de X) est une mesure de probabilité sur \mathbb{R} . Malheureusement il est exceptionnel que l'on dispose d'une expression analytique explicite de la mesure μ permettant par exemple un calcul d'intégrale par quadrature. D'ailleurs lorsqu'une telle situation se présente la connaissance de μ est incomparablement plus riche d'informations que celle de son espérance.

Le plus souvent on procède donc à un calcul approché en se fondant sur la loi forte des grands nombres, dite méthode de Monte Carlo qui assure que :

$$\mathbb{P}(d\omega) - ps \quad \frac{X_1(\omega) + \dots + X_n(\omega)}{n} \xrightarrow{n \rightarrow +\infty} \mathbb{E}X$$

si X_1, \dots, X_n, \dots est une suite de variables aléatoires réelles indépendantes de même loi que X .

Sur un plan pratique la méthode de Monte Carlo consiste à exploiter ce résultat en simulant sur ordinateur (ou en engendrant expérimentalement) le début $X_1(\omega), \dots, X_n(\omega)$ d'une trajectoire $(X_n(\omega))_{n \geq 1}$ de la suite $(X_n)_{n \geq 1}$ puis à assimiler $\mathbb{E}X$ à $\frac{X_1(\omega) + \dots + X_n(\omega)}{n}$ lorsque n est "grand" (i.e. quelques milliers, voire plus). Cela revient à poser :

$$\mathbb{E}X = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(U_{(k-1)d+1}, U_{(k-1)d+2}, \dots, U_{kd}),$$

une fois que X a été simulée sous la forme $X = f(U_1, U_2, \dots, U_d)$. Il est clair que, lorsque la simulation est faite, la programmation de la méthode est immédiate, en plus elle s'applique sans hypothèse de régularité sur f : f est simplement supposée borélienne intégrable sur $[0, 1]^d$.

1.3 Vitesse de convergence de la méthode de Monte Carlo.

Par les inégalités de Bienaymé-Tchebichev, de Chernov ou le théorème de la limite centrale la méthode de Monte Carlo dispose d'un critère d'arrêt probabiliste. C'est à dire que, pour un seuil d'erreur admissible ε donné, on peut a priori décider du nombre de simulations qui rendra la probabilité que ce seuil ne soit pas respecté inférieure à une valeur fixée. Par exemple, si X est de carré intégrable, on utilise l'inégalité de Bienaymé-Tchebichev et on obtient :

$$\mathbb{P}\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X)\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}, \quad (1.1)$$

où $\sigma^2 = \text{Var}(X_1)$. L'utilisation du théorème de la limite centrale avec reste, dit théorème de Berry Essen (cf.[16] et [50]), dans le cas où X admet un moment d'ordre 3, donne une majoration plus fine. En effet ce théorème assure qu'un seuil de précision $\varepsilon > 0$ étant fixé :

$$\mathbb{P}\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}X\right| \geq \varepsilon\right\} \leq \mathbb{P}\left\{|\mathcal{N}(0;1)| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\right\} + \frac{C\rho}{\sigma^3\sqrt{n}(1+\varepsilon\sqrt{n})} \sim \frac{C\rho}{\sigma^3\varepsilon n} \quad (1.2)$$

où $\sigma^2 = \text{Var}(X_1)$, $\rho = E|X_1|^3$ et $\mathcal{N}(0;1)$ désigne une variable aléatoire réelle de loi normale centrée réduite. Lorsque X a des moments exponentiels ($\mathbb{E}e^{b|X|} < \infty$), on utilise un critère plus précis, dit de Chernov, à savoir que :

$$\mathbb{P}\left\{\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}(X)\right| \geq \varepsilon\right\} \leq (\varphi(\varepsilon))^n + (\psi(\varepsilon))^n, \quad (1.3)$$

avec $\varphi(\varepsilon) = \inf_{\gamma \in [0,b]} \mathbb{E}(e^{\gamma(X-\mathbb{E}X)-\gamma\varepsilon})$ et $\psi(\varepsilon) = \inf_{\gamma \in [0,b]} \mathbb{E}(e^{\gamma(\mathbb{E}X-X)-\gamma\varepsilon})$.

D'autre part, la loi du logarithme itéré (cf.[50]) indique, elle, que si $X_1, X_2, \dots, X_n, \dots$ est une suite de variables aléatoires réelles indépendantes de même loi et de carré intégrables, alors :

$$\limsup_{n \rightarrow \infty} \frac{X_1 + \dots + X_n - n\mathbb{E}X}{\sqrt{2n \log \log n}} = \sigma,$$

et

$$\liminf_{n \rightarrow \infty} \frac{X_1 + \dots + X_n - n\mathbb{E}X}{\sqrt{2n \log \log n}} = -\sigma.$$

Ainsi, si $\sigma^2 \neq 0$, autrement dit si X n'est pas déterministe, la quantité $\frac{X_1 + \dots + X_n - n\mathbb{E}X}{\sqrt{2n \log \log n}}$ oscille indéfiniment entre un voisinage de σ et un voisinage de $-\sigma$. On peut dire aussi que : $\mathbb{P}(d\omega) - p.s. \quad \forall \varepsilon > 0 \quad \exists n_{(\omega)}^*$ tel que :

$$\forall n \geq n_{(\omega)}^* \quad \left| \frac{X_1 + \dots + X_n}{n} - \mathbb{E}X \right| \leq (1 + \varepsilon)\sigma \sqrt{\frac{\log \log n}{2n}},$$

mais il ne faut pas attendre de cette méthode une précision meilleur que $\sigma(X) \sqrt{\frac{\log \log n}{2n}}$.

Il convient aussi de remarquer que la vitesse de convergence de la méthode de Monte Carlo ne dépend pas de la simulation de X , autrement dit elle ne dépend pas de la fonction

$f(U_1, \dots, U_d, \dots)$ choisie. En effet, l'erreur asymptotique ne fait intervenir que l'écart-type pour l'inégalité de Bienaymé-Tchebichev et le logarithme-itéré, le moment d'ordre 3 pour le théorème de Berry-Essen, des moments exponentiels pour l'inégalité de Chernov. Dans la suite on comparera la méthode du décalage à la méthode de Monte Carlo dont les résultats ci-dessus conduisent à affirmer que la vitesse de convergence "tourne autour" de $O(\frac{1}{\sqrt{N}})$.

Chapitre 2

Présentation de la méthode du décalage

L'usage du théorème ergodique de Birkhoff pour le calcul d'espérances mathématiques n'est pas très fréquent en simulation. Cependant, dans le cas de certaines transformations, et à cause des particularités de leur implémentation informatique, la programmation de ces algorithmes se trouve être plus efficace. Dans ce chapitre on rappellera quelques résultats de théorie ergodique et on présentera une méthode aléatoire, dite de décalage ou de shift, pour le calcul d'espérance en dimension grande ou infinie (voir [12], [1] et [2]). Cette méthode est basée sur le théorème de Birkhoff.

2.1 Cadre mathématique de la méthode.

Désignons par $(X, \mathcal{A}, \lambda)$ un espace de probabilité et par τ une transformation mesurable de (X, \mathcal{A}) qui préserve la mesure λ (c'est à dire $\lambda(\tau^{-1}A) = \lambda(A)$ pour tout $A \in \mathcal{A}$). On rappelle qu'un ensemble $A \in \mathcal{A}$ est dit τ -invariant si $\tau^{-1}(A) = A$.

Définition 1 Une transformation τ est dite ergodique si les parties mesurables τ -invariantes de \mathcal{A} sont négligeables ou de complémentaires négligeables.

En 1931 Birkhoff a démontré que pour toute fonction f de X , à valeur réelle et λ -intégrable, la limite de $\frac{f(x) + f(\tau x) + \dots + f(\tau^{n-1}x)}{n}$ existe λ -presque sûrement. Il est facile de voir que si la limite est constante elle est égale à l'intégrale de f . On retrouve ainsi le théorème de Birkhoff pour les transformations ergodiques.

Théorème 1 *Birkhoff.*

Soit $(X, \mathcal{A}, \lambda)$ un espace de probabilité et $\tau : X \rightarrow X$ une transformation ergodique alors pour toute fonction $f \in L^1(X, \mathcal{A})$ on a :

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ \tau^k \longrightarrow \int_X f d\lambda \quad \lambda - p.s..$$

Preuve : Voir Krengel [32].

Dans la suite, on rappellera la définition d'une transformation mélangeante et on donnera un critère simple de mélange.

Définition 2 Une transformation mesurable τ d'un espace de probabilité $(X, \mathcal{A}, \lambda)$ est dite *mélangeante* si pour tout $f \in L^2$, la suite $(f \circ \tau^n)$ converge faiblement dans L^2 vers la fonction constante $\int_X f d\lambda$, c'est à dire si pour tout $g \in L^2$

$$\int_X f \circ \tau^n g d\lambda \longrightarrow_{n \rightarrow \infty} \int_X f d\lambda \int_X g d\lambda$$

Manifestement, une transformation mélangeante est ergodique. En effet, en prenant f et g égale à l'indicatrice d'un ensemble mesurable A , on a :

$$\forall A \in (X, \mathcal{A}, \lambda) \quad \lambda(\tau^{-n}(A) \cap A) \longrightarrow_{n \rightarrow \infty} (\lambda(A))^2.$$

Par conséquent, si $\tau^{-1}(A) = A$ alors $\lambda(A) = (\lambda(A))^2$ et donc A est négligeable ou de complémentaire négligeable. La réciproque est fausse. L'introduction de la convergence faible dans la définition précédente est essentielle. En effet la suite $(f \circ \tau^n, n \in \mathbb{N})$ ne peut converger fortement dans L^2 pour un $f \in L^2$ que dans le cas trivial où $f \circ \tau = f$ parce que pour tout $f \in L^2$: $\|f \circ \tau - f\| = \|f \circ \tau^{n+1} - f \circ \tau^n\|$ quel que soit $n \in \mathbb{N}$. Le lemme suivant fournit un critère simple de mélange ; il sera utilisé par la suite pour montrer que la transformation du shift est mélangeante.

Lemme 1 Pour qu'une transformation mesurable τ de $(X, \mathcal{A}, \lambda)$ soit mélangeante il suffit que :

$$\int_X f \circ \tau^n f d\lambda \longrightarrow_{n \rightarrow \infty} \left(\int_X f d\lambda \right)^2$$

pour tout f d'un ensemble générateur de L^2

Preuve : Evident par polarisation et approximation.

Comme nous l'avons dit au premier chapitre, il est très rare, sauf cas simple, que la simulation d'une variable aléatoire réelle se réduise effectivement à une simulation en dimension 1. Cependant, le plus souvent, par simulation d'un certain nombre de variables aléatoires intermédiaires on arrive à définir un algorithme de simulation de la variable aléatoire X . Ainsi, on obtient une représentation de X sur $[0, 1]^{\mathbb{N}}$ ou plus précisément une représentation sur $[0, 1]^{(\mathbb{N})} = \cup_{n \geq 1} [0, 1]^n$. Il est donc intéressant d'étudier un algorithme qui calcule l'intégrale d'une fonctionnelle sur $[0, 1]^{\mathbb{N}}$. Considérons $[0, 1]^{\mathbb{N}}$, l'espace produit de l'intervalle $[0, 1]$, muni de la mesure de Lebesgue produit $\lambda = dx^{\otimes \mathbb{N}}$ et notons par θ la fonction de décalage, appelée aussi opérateur de shift, définie sur $[0, 1]^{\mathbb{N}}$ par

$$\theta(U_1, U_2, \dots, U_k, \dots) = (U_2, U_3, \dots, U_{k+1}, \dots).$$

Il est clair que $\theta(\lambda) = \lambda$, où $\theta(\lambda)$ est la mesure image. Dans le lemme suivant on rappelle un résultat classique à savoir que θ est non seulement ergodique mais mélangeante.

Lemme 2 θ est une transformation mélangeante qui préserve la mesure λ .

Preuve : Pour établir ce résultat il suffit, d'après le lemme précédent, de démontrer que :

$$\forall A \in \mathcal{B}([0, 1]^{\otimes N}) \quad \text{on a} \quad \lambda(\theta^{-n}(A) \cap A) \xrightarrow{n \rightarrow \infty} \lambda(A)\lambda(A).$$

Il suffit de vérifier cette propriété sur la classe des ensembles cylindriques. Soit A un ensemble $\sigma(U_k, \dots, U_l)$ -mesurable on a $\theta^{-n}(A)$ est $\sigma(U_{k+n}, \dots, U_{l+n})$ -mesurable. Par suite, pour n assez grand on a $\theta^{-n}(A)$ et A sont indépendants, par conséquent $\lambda(\theta^{-n}(A) \cap A) = \lambda(\theta^{-n}(A))\lambda(A) = (\lambda(A))^2$. Ceci achève la démonstration. ■

Le système dynamique $([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$ est donc en particulier ergodique. Par conséquent, d'après le théorème ergodique ponctuel de Birkhoff on a :

$$\frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{N \rightarrow \infty} \mathbb{E}F \quad \lambda - p.s., \quad (2.1)$$

pour toute fonctionnelle F λ -intégrable sur $[0, 1]^N$. Donc une simulation de l'espérance mathématique d'une variable aléatoire réelle représentée sur $[0, 1]^N$ par $X = F(U) = F(U_1, \dots, U_k, \dots)$ consiste à calculer la fonctionnelle F aux points successifs $U, \theta(U), \theta^2(U), \dots$. Ces points sont des suites de points de $[0, 1]$:

$$\begin{aligned} U &= (U_1, U_2, \dots, U_k, \dots) \\ \theta(U) &= (U_2, U_3, \dots, U_{k+1}, \dots) \\ &\vdots \end{aligned}$$

et de prendre la moyenne des termes.

Remarque :

Nous désignerons ici par θ^d les puissances de θ avec $d \in \mathbb{N}^*$. Il convient donc de remarquer que les résultats ci-dessus restent valables pour l'espace $([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta^d)$. En particulier, le système dynamique $([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta^d)$ est ergodique et par suite :

$$\frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^{nd} \xrightarrow{N \rightarrow \infty} \mathbb{E}F \quad \lambda - p.s., \quad (2.2)$$

pour toute fonctionnelle F λ -intégrable sur $[0, 1]^N$.

Dans le prochain paragraphe on expliquera l'intérêt de cette méthode en montrant par un exemple générique le principe du dispositif informatique.

2.2 Principe du dispositif informatique.

La modélisation des phénomènes aléatoires, fait souvent intervenir des processus à temps discret, ainsi les problèmes de transport de particules, la discrétisation des équations différentielles stochastiques, etc. Pour calculer l'espérance d'une fonctionnelle des trajectoires

d'un processus, on est ramené à des intégrations en dimension infinie et on se retrouve dans des situations voisines de celle ci : Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur $[0, 1]^{\mathbb{N}}$ telle que, $\forall n \in \mathbb{N}$, X_n est $\sigma(U_1, \dots, U_n)$ -mesurable et soit T un temps d'entrée du processus $(X_n)_{n \in \mathbb{N}}$, ou de $(X_n, n)_{n \in \mathbb{N}}$, dans un ensemble donné, on a à calculer les quantités suivantes :

- $\mathbb{E}(T)$, $\mathbb{E}(X_T)$, $\mathbb{E}[G(X_T, T)]$, $G : \mathbb{R} \otimes \mathbb{N} \longrightarrow \mathbb{R}$.
- $\mathbb{E} \sum_{k=1}^T C(X_k, X_{k+1})$, où C est une fonction de \mathbb{R}^2 à valeurs réelles, représentant des coûts.
- Ou plus généralement $\mathbb{E}(H)$ avec H une fonctionnelle \mathcal{F}_T -mesurable.

Très souvent ces processus sont des chaînes de Markov. Nous allons décrire, par un exemple générique, comment s'implémente la simulation d'une telle chaîne par la méthode du décalage. Soit X_n une chaîne de Markov définie par :

$$X_0 = x, \quad X_{n+1} = H(X_n, n, h(U_{nd+1}, \dots, U_{(n+1)d})),$$

où les U_n sont des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Ainsi le processus $(X_n)_{n \in \mathbb{N}}$ admet une représentation sur $([0, 1]^{\mathbb{N}}, dx^{\otimes \mathbb{N}})$.

On considère le temps d'arrêt $T = \inf\{n \geq 1; X_n \in A\}$ et les fonctionnelles associées $F = G(X_T, T)$. L'espérance $\mathbb{E}G(X_T, T)$ peut être calculée par le théorème de Birkhoff, puisque

$$\mathbb{E}G(X_T, T) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} G(X_T, T) \circ \theta^{nd} \quad \lambda - p.s., \quad (2.3)$$

dès que $G(X_T, T) \in L^1$. L'implémentation de cet algorithme exploite de façon essentielle la notion de pointeur (ou tout autre procédé équivalent de gestion de piles). Pour appliquer la formule (2.3) nous avons à calculer $G(X_T, T)$ aux points successifs $U, \theta^d(U), \theta^{2d}(U), \dots$. Ces points sont des éléments de $[0, 1]^{\mathbb{N}}$

$$\begin{aligned} U &= (U_1, U_2, \dots, U_d, \dots, U_k, \dots) \\ \theta^d(U) &= (U_{d+1}, U_{d+2}, \dots, U_{2d}, \dots, U_{d+k}, \dots) \\ &\vdots \end{aligned}$$

Alors si le temps d'arrêt T est fini, ce que nous supposons ici, le calcul $G(X_T, T)$ ne dépend que d'un nombre fini de coordonnées, le long de chaque trajectoire $U = (U_1, U_2, \dots, U_k, \dots)$: la suite U étant calculée jusqu'à l'indice $k = dT(U)$ où $T(U)$ est déterminé par le test

$$< test > \begin{cases} X_n \notin A & \Rightarrow n < T(U) \\ X_n \in A \text{ pour la première fois} & \Rightarrow n = T(U) \end{cases}$$

Notons par :

$$Y_n = h(U_{(n-1)d+1}, \dots, U_{nd}).$$

Supposons qu'on ait calculé $G(X_T, T)$ au point $U = (U_1, U_2, \dots)$, et que les calculs des termes intermédiaires, $Y_n = h(U_{(n-1)d+1}, \dots, U_{nd})$, $1 \leq n \leq T(U)$, aient été mis dans des

pointeurs comme indiqué dans la figure 2.1. La simulation de $G(X_T, T)$ sur la trajectoire $\theta^d(U) = (U_{d+1}, U_{d+2}, \dots)$ sera plus simple puisque le calcul intermédiaire des Y_i associés à $\theta^d(U)$ est fait au moins partiellement. En effet, soit on a $T(\theta^d(U)) < T(U)$ auquel cas aucun calcul supplémentaire n'est requis, soit on a simplement à rallonger la chaîne entre $T(U)$ et $T(\theta^d(U))$ auquel cas nous avons la figure 2.2 :

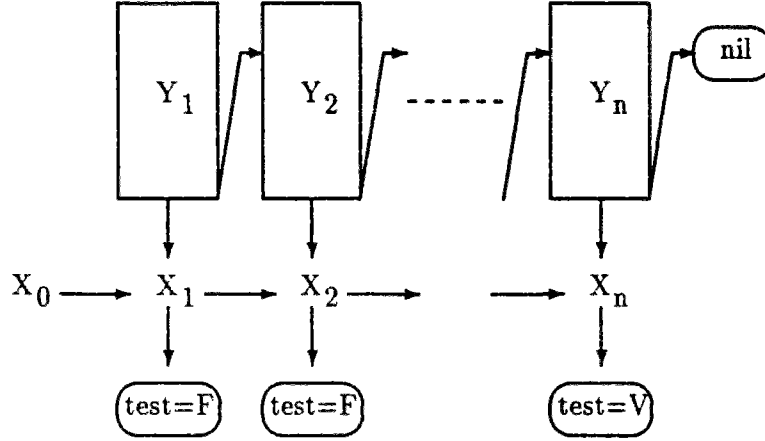


Figure 2.1

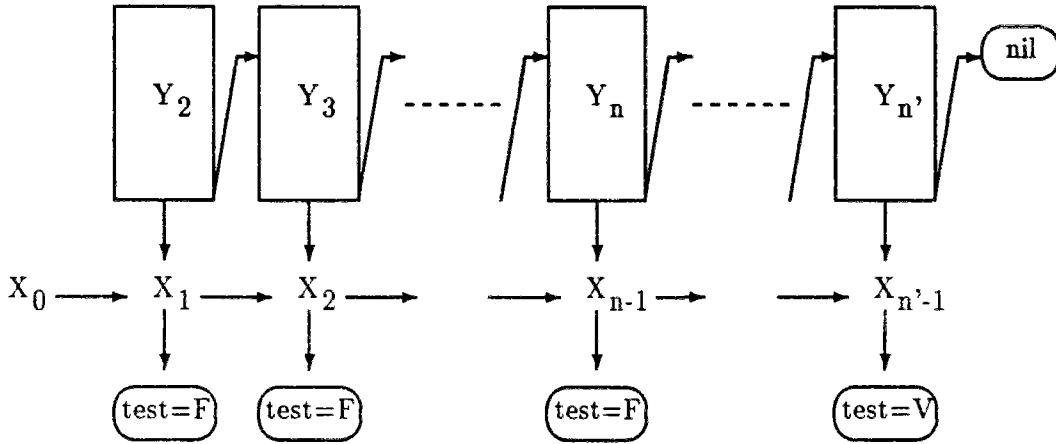


Figure 2.2

Dans ces diagrammes les flèches représentent essentiellement le calcul de :

$$X_{n+1} = H(X_n, n, h(U_{nd+1}, \dots, U_{(n+1)d})) = H(X_n, n, Y_{n+1}).$$

Il est clair que les performances de cette méthode proviennent de la boîte de stockage qui évite de recommencer inutilement les calculs partiels des variables Y_i . Ce stockage doit être fait évidemment lors du rallongement de la suite lorsque ce dernier intervient comme expliqué ci-dessus.

Dans la suite de ce chapitre nous testerons l'efficacité de cette méthode, nous la comparerons à la méthode de Monte Carlo classique et nous mettrons en évidence par des exemples l'économie de temps de calcul, mais aussi de nombres aléatoires qu'elle permet de réaliser.

2.2.1 Quelques résultats numériques sur l'efficacité du shift.

Pour comparer les deux méthodes on a calculé sur plusieurs exemples le temps machine, dit temps C.P.U, pour chaque algorithme. Pour un même rang d'itération N , on remarque que la méthode du décalage est nettement plus rapide que la méthode de Monte Carlo classique. Le gain du temps C.P.U est en moyenne de l'ordre de 90%. Ce gain peut être plus important si l'on prend des situations où le calcul de la fonction h est très long, autrement dit si l'on considère des exemples où la simulation des Y_n est donnée par des méthodes lentes ; par exemple : la méthode du rejet, schéma de résolution d'une équation différentielles stochastique. En outre, cette méthode économise le générateur de nombres aléatoires. En effet, la consommation de nombres aléatoires doit être limitée à une certaine fraction du générateur pour en préserver la qualité. Il convient aussi de remarquer que la méthode de Monte Carlo pour des problèmes de taille équivalente risque souvent de dépasser ce seuil. Tous les résultats numériques donnés dans les exemples suivants sont calculés sur des SUN4/330.

Exemple 1

Une façon de simuler la marche aléatoire réelle de loi normale centrée réduite, issue de l'origine, est de prendre : $X_0 = 0$, $X_{n+1} = X_n + \sqrt{-2 \ln U_{2n+1}} \cos(2\pi U_{2n+2})$, où $(U_n)_{n \in \mathbb{N}}$ est une suite de variable aléatoires indépendantes de loi uniforme sur $[0,1]$. Considérons le temps d'arrêt $T = \inf\{n \geq 1, X_n \notin [-10, 10]\}$. Pour des fonctionnelles différentes, $F = X_T$ et $F = T$, on a les résultats suivants.

- Pour $F = X_T$.

Nombres d'itérations	Temps C.P.U.		Nombres aléatoires	
	Shift	Monte Carlo	Shift	Monte Carlo
1000	1.3	18.0	2876	220232
5000	5.2	93.1	10214	1132330
10000	10.5	199.5	20610	2260466
50000	59.0	916.4	100664	11209942
100000	118.8	2078.1	200108	22440720

Tableau 2.1

- Pour $F = T$.

Nombres d'itérations	Temps C.P.U.		Nombres aléatoires	
	Shift	Monte Carlo	Shift	Monte Carlo
1000	1.3	18.2	2876	220232
5000	5.3	94.3	10214	1132330
10000	10.8	187.2	20610	2260466
50000	60.4	936.4	100664	11209942
100000	123.1	1854.1	200108	22440720

Tableau 2.2

Ces résultats confirment que le gain du temps C.P.U. ainsi que celui en nombres aléatoires dépend essentiellement du temps d'arrêt T et non de la fonctionnelle F . Ceci était d'ailleurs tout à fait prévisible.

Exemple 2

Dans la pratique, les spécialistes utilisent rarement les nombres pseudo-aléatoires fournis par l'ordinateur. Ils les génèrent par une procédure. Ceci explique le fait que, même dans un exemple simple où les variables intermédiaires $(Y_n)_{n \in \mathbb{N}}$ suivent une loi uniforme, la méthode du shift paraisse plus efficace. On a testé ce résultat sur une marche aléatoire réelle de loi uniforme sur $[-0.5, 0.5]$, issue de l'origine. Prenons donc : $X_0 = 0$, et $X_{n+1} = X_n + \frac{2U_{n+1}-1}{2}$, où $(U_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes, de loi uniforme sur $[0, 1]$. Considérons le temps d'arrêt $T = \inf\{n \geq 1, X_n \notin [-5, 1]\}$. Pour les fonctionnelles $F = X_T$ et $F = T$ on a les résultats suivants.

- Pour $F = X_T$.

Nombres d'itérations	Temps C.P.U.		Nombres aléatoires	
	Shift	Monte Carlo	Shift	Monte Carlo
1000	0.270	2.710	1064	69183
5000	1.680	13.80	5208	355140
10000	3.460	27.890	10069	718189
50000	17.680	138.870	50042	3581028
100000	38.160	276.320	100298	7122491

Tableau 2.3

- Pour $F = T$.

Nombres d'itérations	Temps C.P.U.		Nombres aléatoires	
	Shift	Monte Carlo	Shift	Monte Carlo
1000	0.280	2.720	1064	69183
5000	1.680	13.770	5208	355140
10000	3.510	27.830	10069	718189
50000	18.900	138.670	50042	3581028
100000	38.280	269.630	100298	7122491

Tableau 2.4

Exemple 3

Nous avons utilisé la méthode du shift pour une chaîne de Markov à 2 dimensions. Dans une étude faite par O. Faure([15]) sur le temps de sortie d'un disque pour un Brownien plan $(B_t)_{t \geq 0}$, il a été confronté à une situation typique où la dimension de la simulation est très grande. Si l'on utilise la méthode de Monte Carlo, avec un pas de discrétisation trop fin, alors le temps de calcul devient très important et le nombre trop élevé d'appels à la fonction

random peut engendrer des erreurs de calculs. Soit par arrondi soit, pire, par épuisement du générateur. Pour contourner cette difficulté nous avons utilisé la méthode du shift. Ceci nous a permis de confirmer l'intérêt de cette méthode, à savoir l'économie de nombre d'appels à la fonction random et celui du temps de calcul.

Plaçons nous dans \mathbb{R}^2 . Soit C le disque de centre 0 et de rayon 1.

On désigne par

$$T = \inf\{t > 0, |B_t| \notin C\}$$

le temps de sortie du disque C . La simulation du Brownien B_t est donné par la simulation d'une chaîne de Markov définie par :

$$X_0 = 0, \quad X_{k+1} = X_k + Y_{k+1}$$

avec Y_{k+1} est une suite de variables aléatoires dans \mathbb{R}^2 indépendantes, de loi gaussiennes centrées et de variance Δt . Le paramètre Δt est appelé le pas de discrétisation. Une simulation de la suite $(Y_k)_{k \in \mathbb{N}}$ est donnée par :

$$\begin{aligned} Y_k^1 &= \sqrt{-2\Delta t \ln U_{2k}} \cos(2\pi U_{2k+1}) \\ Y_k^2 &= \sqrt{-2\Delta t \ln U_{2k}} \sin(2\pi U_{2k+1}) \end{aligned}$$

où Y_k^1 est l'abscisse de Y_k , Y_k^2 est l'ordonnée de Y_k et $(U_n)_{n \in \mathbb{N}}$ est une suite de variable aléatoires indépendantes de loi uniforme sur $[0,1]$. Le temps de sortie T est pris comme le premier temps de sortie, $T^{\Delta t}$, de la chaîne de Markov approchante. On vérifie classiquement à l'aide de résultats de convergence fonctionnelle (voir[15]) que :

$$\lim_{\Delta t \rightarrow 0} |\mathbb{E}(T^{\Delta t}) - \mathbb{E}(T)| = 0.$$

La valeur théorique $\mathbb{E}T = \frac{1}{2}$ est prise comme critère de comparaison des deux méthodes. Tout d'abord nous allons montrer les difficultés numériques rencontrées par la méthode de Monte Carlo. Nous donnons dans le tableau suivant, pour quelques valeurs du paramètre Δt des statistiques sur le nombre d'appels à la fonction random et le temps mis par la machine pour faire les calculs sur une seule trajectoire :

- Temps C.P.U-moyen : le temps moyen de calcul d'une simulation en C.P.U. Plus précisément, on a pris la moyenne des temps de calcul sur 10 000 simulations.
- Nbres aléatoires-moyen, le nombre moyen d'appels à la fonction random pour simuler une trajectoire de sortie.

Le temps moyen, τ -moyen, dépend bien sûr de la machine utilisé (ici un SUN4/330).

Δt	0.05	0.01	0.005	0.001	0.0005	0.0001	0.00001
Temps C.P.U-moyen	0.0048	0.0206	0.0396	0.1841	0.3691	1.6611	17.7
Nbres aléatoires-moyen	26	112	216	1000	2000	9000	80 000

Tableau 2.5

Ceci permet de justifier l'utilisation de la méthode du shift puisqu'on simule en dimension grande. En effet, la taille de simulation est équivalente au nombre moyen d'appels à la fonction random.

Nous calculons ensuite l'espérance par Monte Carlo sur 10 000 simulations suivant différentes valeurs du paramètre de discrétisation Δt . Nous donnons aussi quelques résultats de performance pouvant servir d'éléments de comparaison :

- Temps C.P.U.-total : le temps de calcul des 10 000 simulations en temps C.P.U.
- Nbres aléatoires-total : le nombre d'appels à la fonction random pour simuler les 10 000 trajectoires de sortie.

Δt	0.05	0.01	0.005	0.001	0.0005	0.0001
$\mathbb{E}(T - \frac{1}{2})$	0.1588	0.0608	0.0420	0.0209	0.0124	-0.0042
Temps C.P.U.-total	48	206	396	1840	3690	16 611
Nbres aléatoires-total	260 000	1 120 000	2 160 000	10^7	$2 \cdot 10^7$	$9 \cdot 10^8$

Tableau 2.6

Nous calculons maintenant l'espérance par la méthode du shift sur $N = 10\,000$ itérations et suivant les mêmes valeurs du paramètre de discrétisation Δt .

Δt	0.05	0.01	0.005	0.001	0.0005	0.0001
$\mathbb{E}(T - \frac{1}{2})$	0.1383	0.0560	0.0485	0.0215	0.0134	-0.0638
τ -total	7	20	35	170	300	900
Nbres aléatoires-total	20 020	20 070	20 113	20 500	21 000	29 000

Tableau 2.7

Dans cet exemple, on remarque que l'erreur théorique est pratiquement du même ordre. Cependant, le gain du temps C.P.U et celui des nombres aléatoires est très important. En particulier, lorsque le pas de discrétisation Δt est très fin. Une étude plus précise sur la résolution des équations différentielles stochastiques par la méthode du shift sera faite au chapitre 6.

2.3 Les résultats existants sur la vitesse de convergence.

L'utilisation de la méthode du shift nécessite des estimations de la vitesse de convergence, tels que la loi du logarithme itéré et le théorème de la limite centrale. Malheureusement, le théorème ergodique ponctuel de Birkhoff ne dispose pas d'un tel résultat. Des études sur le comportement asymptotique dans le théorème ergodique (voir [33], [22]), montrent que la vitesse de convergence peut être arbitrairement lente ou proche de $0(\frac{1}{N})$. En effet, si on considère le système dynamique $([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$, on a :

D'une part : pour toute suite $(\alpha_n)_{n \in \mathbb{N}}, \alpha_n > 0, \alpha_n \rightarrow 0$, il existe une fonction continue F de $[0, 1]^{\mathbb{N}}$ à valeurs réelles telle que

$$\frac{1}{\alpha_N} \left(\left| \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n - \mathbb{E}F \right| \right) \rightarrow \infty, \quad \lambda - p.s..$$

D'autre part : pour toute suite $(c_n)_{n \in \mathbb{N}}, c_n > 0$, croissant vers l'infini, avec $c_1 \geq 2$, il existe un ensemble mesurable A avec $P(A) = \frac{1}{2}$, pour lequel

$$\forall N \quad \left| \frac{1}{N} \sum_{n=0}^{N-1} 1_A \circ \theta^n - \frac{1}{2} \right| \leq \frac{c_N}{N}.$$

Par ailleurs, soit $(E, || ||)$ un espace de Banach et U une transformation linéaire de E dans E . On dit que $a \in E$ est un U -cobord si et seulement si $a \in (I - U)(E)$, autrement dit s'il existe $b \in E$ tel que $a = b - U(b)$. Il est clair que si a est un U -cobord alors la suite $\frac{a + U(a) + \dots + U^{n-1}(a)}{n}$ converge en $O(\frac{1}{n})$. Les cobords ont été étudiés dans le cas de certaines transformations ergodiques par exemple : voir [34] pour les suites $\{n\alpha\}$ et [39] pour les transformations de Kakutani. On montre que, pour ces transformations particulières, il existe des ensembles A non triviaux tels que les 1_A sont des cobords.

Ces fonctions, pour celles actuellement connues comme étant lentement convergentes, sont construites avec le lemme de Rohlin-Halmos. Par la nature abstraite de ce lemme, l'intérêt de ces résultats est essentiellement théorique. En effet, il se trouve que les fonctionnelles courantes en simulation ne présentent pas la pathologie des cas lentement convergent du théorème ergodique. Ainsi, notre objectif est d'obtenir des estimations de vitesse de convergence pour des classes de fonctions couramment employées en pratique.

Chapitre 3

Vitesse de convergence presque sûre.

Le but de ce chapitre est de déterminer un ordre de convergence de $\frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n$ vers $\mathbb{E}F$ pour certaines classes raisonnablement générales de fonctionnelles F couramment employées en simulation. Ces résultats s'obtiennent à partir d'un théorème de Gál et Koksma (voir [18]). Dans le premier paragraphe nous rappelons l'énoncé de ce théorème, sa démonstration dans un cadre probabiliste et sa formulation dans le cadre de l'opérateur de décalage. Le second paragraphe est consacré à la proposition théorique principale sur la vitesse de convergence dans le théorème ponctuel ergodique pour l'opérateur de décalage, à sa démonstration et à une première application. Dans le troisième paragraphe, le plus utile pour les applications, on applique cette proposition à la classe des fonctionnelles F de carré intégrable et mesurables par rapport à la tribu \mathcal{F}_T des événements antérieurs à un temps d'arrêt T . On montre notamment que si T admet un moment d'ordre $p > 2$ et si F est une fonction de L^2 , \mathcal{F}_T -mesurable, alors F vérifie "une loi du logarithme itéré faible", à savoir que :

$$\forall \varepsilon > 0 \quad \left| \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n - \mathbb{E}F \right| = o(N^{-\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}) \quad \lambda - p.s.. \quad (3.1)$$

Le quatrième paragraphe illustre par des simulations numériques les résultats théoriques obtenus sur les fonctionnelles d'arrêt. En outre, nous avons vérifié la loi du logarithme itéré habituelle par des simulations numériques. Nous y reviendrons au chapitre 5 pour donner les résultats théoriques.

3.1 Théorème de Gál et Koksma et vitesse de convergence dans le théorème de Birkhoff.

Tout d'abord nous allons donner quelques définitions.

Soit (X, μ) un espace mesuré avec μ une mesure non nécessairement finie. On considère une famille de fonctions $(S(M, N, \cdot))_{M, N \in \mathbb{N}}$, appartenant à $L^p(X, \mu)$ et telles que $S(M, N, x) \geq 0$ pour tout M et N appartenant à \mathbb{N} et $x \in X$. On suppose enfin que $S(M, 0, \cdot) = 0$ pour tout $M \in \mathbb{N}$ et que $S(M, N, \cdot)$ vérifie l'inégalité triangulaire

$$S(M, N, \cdot) \leq S(M, N', \cdot) + S(M + N', N - N', \cdot) \quad (3.2)$$

pour tout $M, N, N' \in \mathbb{N}$ et $0 \leq N' \leq N$. Typiquement, $S(M, N, \cdot)$ est la valeur absolue d'une sommation partielle $S(M, N, x) = \left| \sum_{k=M+1}^N f_k(x) \right|$. A partir de ces définitions, on démontre le résultat suivant.

Théorème 2 *On suppose que $p \geq 1$.*

$$\text{Si} \quad \int_X S(M, N, x)^p d\mu = O(\Psi(N)) \quad \text{uniformément en } M, \quad (3.3)$$

où $\frac{\Psi(N)}{N}$ est une fonction croissante au sens large, alors pour tout $\varepsilon > 0$ on a :

$$S(0, N, x) = o\left((\Psi(N) (\log(N))^{p+1+\varepsilon})^{\frac{1}{p}}\right) \quad \mu - p.s. \text{ dans } X.$$

Preuve : Ce résultat a été démontré dans l'article de Gál et Koksma "Sur l'ordre de grandeur des fonctions sommables", sous l'hypothèse que X est une partie mesurable d'un espace euclidien, voir référence [18], théorème 3, page 646 et la définition, page 640. ■

Le but de l'article était d'appliquer ce résultat à la somme partielle d'une série, $(f_N(x))_{N \in \mathbb{N}}$, non nécessairement convergente. En effet, si on pose $S(M, N, x) = \left| \sum_{k=M+1}^N f_k(x) \right|$ où les fonctions $f_N(x)$ sont à p -ième puissance sommables, le théorème donne une estimation pour la somme $f_1(x) + f_2(x) + \dots + f_N(x)$ qui est valable presque partout.

Toutes ces remarques nous ont conduit à reprendre la démonstration de ce théorème avec des notations et des arguments de théorie de la mesure, essentiellement le lemme de Borel-Cantelli et l'inégalité de Bienaymé-Tchebichev, de façon à mieux mettre en valeur les idées sous-jacentes. A l'inverse, l'hypothèse " X est une partie mesurable d'un espace euclidien" qui figurait dans l'énoncé "historique" est en fait inutile. Comme nous ne ferons usage de ce théorème que dans un cadre probabiliste nous nous placerons dans la démonstration sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ avec $\mathbb{P}(\Omega) = 1$. Cette hypothèse n'est cependant pas nécessaire. On pourra alors, si l'on considère une suite de variables aléatoires $(X_i)_{i \in \mathbb{N}}$ sur un espace probabilisé quelconque $(\Omega, \mathcal{A}, \mathbb{P})$ avoir une estimation du comportement de la somme

$$|X_1(\omega) + X_2(\omega) + \dots + X_N(\omega)|,$$

valable $\mathbb{P}(d\omega)$ -presque sûrement.

Théorème 3 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires appartenant à L^p , $p \geq 1$, vérifiant*

$$\mathbb{E}|X_{M+1} + X_{M+2} + \dots + X_{M+N}|^p = O(\Psi(N)) \quad (3.4)$$

uniformément en M où $\frac{\Psi(N)}{N}$ est une fonction croissante au sens large. Alors pour tout $\varepsilon > 0$, on a :

$$X_1(\omega) + X_2(\omega) + \dots + X_N(\omega) = o\left((\Psi(N) (\log(N))^{p+1+\varepsilon})^{\frac{1}{p}}\right) \quad \mathbb{P}(d(\omega)) - p.s.$$

Preuve : Soit $a > 0$ et $\varepsilon > 0$ fixés. Pour tout $k \geq 1$, on pose :

$$A_k = \bigcup_{N \in [2^k, 2^{k+1}[} \left\{ \left| \sum_{i=1}^N X_i \right| \geq a(\Psi(2^k) k^{p+1+\varepsilon})^{\frac{1}{p}} \right\}.$$

Le problème se ramène à montrer que $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$. En effet, d'après le lemme de

Borel Cantelli, on obtient alors que pour tout $a > 0$ $\mathbb{P} \left(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k \right) = 0$, c'est à dire

$\mathbb{P} \left(\bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k^c \right) = 1$, où A_k^c désigne le complémentaire de l'événement A_k . Autrement dit :

$\forall a > 0$ $\mathbb{P}(d\omega) - p.s.$ il existe $n \in \mathbb{N}$ tel que $\forall k \geq n$, $\forall N \in [2^k, 2^{k+1}[$ on a

$$\left| \sum_{i=1}^N X_i \right| < a(\Psi(2^k) k^{p+1+\varepsilon})^{\frac{1}{p}}.$$

Il résulte de la monotonie de la fonction $\frac{\Psi(N)}{N}$ et celle de la fonction logarithme que

$\Psi(2^k) \leq \Psi(N) \frac{2^k}{N} \leq \Psi(N)$ et que $k \leq \frac{\log(N)}{\log(2)}$. Ceci montre que : $\forall a > 0$ $\mathbb{P}(d\omega) - p.s.$ il

existe $n \in \mathbb{N}$ tel que $\forall N \geq 2^n$ $\left| \sum_{i=1}^N X_i \right| < c(\Psi(N) (\log(N))^{p+1+\varepsilon})^{\frac{1}{p}}$, avec $c = \frac{a}{(\log 2)^{\frac{p+1+\varepsilon}{p}}}$.

Soit encore

$$\forall \varepsilon > 0 \quad \left| \sum_{i=1}^N X_i \right| = O \left((\Psi(N) (\log(N))^{p+1+\varepsilon})^{\frac{1}{p}} \right) \quad \mathbb{P}(d\omega) - p.s.$$

Du fait que l'estimation est donnée pour $\varepsilon > 0$, on peut conclure que :

$$\forall \varepsilon > 0 \quad \left| \sum_{i=1}^N X_i \right| = o \left((\Psi(N) (\log(N))^{p+1+\varepsilon})^{\frac{1}{p}} \right) \quad \mathbb{P}(d\omega) - p.s.$$

Il reste donc à prouver que, $\forall a > 0$ on a $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$. Pour ceci, on considère le développement dyadique de $N \in \{2^k, \dots, 2^{k+1} - 1\}$

$$N = 2^k + \varepsilon_{k-1} 2^{k-1} + \dots + \varepsilon_1 2 + \varepsilon_0 \quad (3.5)$$

où $\varepsilon_i \in \{0, 1\}$ pour $i = 0, 1, \dots, k-1$. D'après l'inégalité triangulaire on obtient de proche en proche, que

$$\begin{aligned} \left| \sum_{i=1}^N X_i \right| &\leq \left| \sum_{i=1}^{2^k} X_i \right| + \left| \sum_{i=2^k+1}^{2^k+\varepsilon_{k-1}2^{k-1}} X_i \right| + \dots \\ &\quad \dots + \left| \sum_{i=2^k+\varepsilon_{k-1}2^{k-1}+\dots+\varepsilon_1 2+\varepsilon_0}^{2^k+\varepsilon_{k-1}2^{k-1}+\dots+\varepsilon_1 2+\varepsilon_0+1} X_i \right|. \end{aligned}$$

Par convention, on prendra la somme égale à 0 si l'exposant en haut est plus petit que l'exposant en bas. On pose :

$$S(M, N, \omega) = |X_{M+1}(\omega) + X_{M+2}(\omega) + \cdots + X_{M+N}(\omega)|,$$

pour tout M et $N \in \mathbb{N}^*$ et $S(M, 0, \omega) = 0$. L'inégalité précédente s'écrit :

$$|\sum_{i=1}^N X_i| \leq S(0, 2^k, \cdot) + \sum_{j=1}^k S(\lambda_j, \varepsilon_{j-1} 2^{j-1}, \cdot).$$

avec $\lambda_j = \varepsilon_j 2^j + \cdots + 2^k$. Soit encore

$$|\sum_{i=1}^N X_i| \leq S(0, 2^k, \cdot) + \sum_{j=1}^k \varepsilon_{j-1} S(\lambda_j, 2^{j-1}, \cdot).$$

En majorant ε_{j-1} par 1, on obtient le résultat suivant :

$$|\sum_{i=1}^N X_i| \leq S(0, 2^k, \cdot) + \sum_{j=1}^k S(\lambda_j, 2^{j-1}, \cdot). \quad (3.6)$$

On remarque que cette majoration dépend de N . Cependant, on est en train de chercher un majorant du $\sup_{N \in [2^k, 2^{k+1}[} |\sum_{i=1}^N X_i|$.

• si $p=1$:

On majore $\sup_{N \in [2^k, 2^{k+1}[} |\sum_{i=1}^N X_i|$ par

$$S(0, 2^k, \cdot) + \sum_{\lambda_j \text{ possible}} S(\lambda_j, 2^{j-1}, \cdot).$$

Or $\lambda_j = 2^j \mu_j + 2^k$ avec $0 \leq \mu_j \leq 2^{k-j} - 1$. Par conséquent :

$$|\sum_{i=1}^N X_i| \leq S(0, 2^k, \cdot) + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} S(2^j \mu_j + 2^k, 2^{j-1}, \cdot) \quad (3.7)$$

et donc

$$\mathbb{P}(A_k) \leq \mathbb{P} \left\{ S(0, 2^k, \cdot) + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} S(2^j \mu_j + 2^k, 2^{j-1}, \cdot) \geq a(\Psi(2^k)(k)^{2+\varepsilon}) \right\}.$$

Il résulte de l'inégalité de Bienaymé-Tchebichev que :

$$\begin{aligned} \mathbb{P}(A_k) &\leq \frac{1}{a\Psi(2^k)k^{2+\varepsilon}} \mathbb{E} \left(S(0, 2^k, \cdot) + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} S(2^j \mu_j + 2^k, 2^{j-1}, \cdot) \right) \\ &\leq \frac{C}{a\Psi(2^k)k^{2+\varepsilon}} \left(\Psi(2^k) + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} \Psi(2^j) \right), \end{aligned}$$

avec C la constante sous-jacente du "O" de l'erreur. Soit encore

$$\mathbb{P}(A_k) \leq \frac{C}{ak^{2+\varepsilon}} \left(1 + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} \frac{\Psi(2^j)}{\Psi(2^k)} \right).$$

Par la monotonie de la fonction $\frac{\Psi(N)}{N}$, on déduit que

$$\begin{aligned} \mathbb{P}(A_k) &\leq \frac{C}{a k^{2+\varepsilon}} \left(1 + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} 2^{j-k} \right) \\ &\leq \frac{C(k+1)}{a k^{2+\varepsilon}} \end{aligned}$$

Ceci achève la démonstration dans le cas $p = 1$.

• si $p > 1$:

Soit $q > 1$ tel que $\frac{1}{p} + \frac{1}{q} = 1$. En appliquant l'inégalité de Hölder au niveau de l'équation (3.6), on obtient :

$$\begin{aligned} \left| \sum_{i=1}^N X_i \right| &\leq \left(1 + \sum_{j=1}^k 1 \right)^{\frac{1}{q}} \left(S(0, 2^k, \cdot)^p + \sum_{j=1}^k S(\lambda_j, 2^{j-1}, \cdot)^p \right)^{\frac{1}{p}} \\ &\leq (1+k)^{\frac{1}{q}} \left(S(0, 2^k, \cdot)^p + \sum_{j=1}^k S(\lambda_j, 2^{j-1}, \cdot)^p \right)^{\frac{1}{p}} \end{aligned}$$

On majore alors $\sup_{N \in [2^k, 2^{k+1}[} \left| \sum_{i=1}^N X_i \right|$ par

$$(1+k)^{\frac{1}{q}} \left(S(0, 2^k, \cdot)^p + \sum_{\lambda, \text{possible}} S(\lambda_j, 2^{j-1}, \cdot)^p \right)^{\frac{1}{p}}$$

Comme dans le cas $p=1$, on obtient

$$\left| \sum_{i=1}^N X_i \right| \leq (1+k)^{\frac{1}{q}} \left(S(0, 2^k, \cdot)^p + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} S(\lambda_j, 2^{j-1}, \cdot)^p \right)^{\frac{1}{p}}.$$

Puis :

$$\begin{aligned} \mathbb{P}(A_k) &\leq \mathbb{P} \left\{ S(0, 2^k, \cdot)^p + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} S(\lambda_j, 2^{j-1}, \cdot)^p \geq \frac{a^p (\Psi(2^k) k^{p+1+\varepsilon})}{(1+k)^{\frac{p}{q}}} \right\} \\ &\leq \frac{(1+k)^{\frac{p}{q}}}{a^p \Psi(2^k) k^{p+1+\varepsilon}} \mathbb{E} \left(S(0, 2^k, \cdot)^p + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} S(2^j \mu_j + 2^k, 2^{j-1}, \cdot)^p \right) \\ &\leq \frac{C(1+k)^{\frac{p}{q}}}{a^p \Psi(2^k) k^{p+1+\varepsilon}} \left(\Psi(2^k) + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} \Psi(2^j) \right) \\ &\leq \frac{C(1+k)^{\frac{p}{q}}}{a k^{p+1+\varepsilon}} \left(1 + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} \frac{\Psi(2^j)}{\Psi(2^k)} \right) \end{aligned}$$

Par la monotonie de la fonction $\frac{\Psi(N)}{N}$, on en déduit que

$$\begin{aligned} \mathbb{P}(A_k) &\leq \frac{C(1+k)^{\frac{p}{q}}}{a^p k^{p+1+\epsilon}} \left(1 + \sum_{j=1}^k \sum_{\mu_j=0}^{2^{k-j}-1} 2^{j-k} \right) \\ &\leq \frac{C(1+k)^{1+\frac{p}{q}}}{a^p k^{p+1+\epsilon}}. \end{aligned}$$

Du fait que $1 + \frac{p}{q} = p$, la dernière inégalité s'écrit : $\mathbb{P}(A_k) \leq \frac{C(1+k)^p}{a^p k^{p+1+\epsilon}}$.

Ainsi la démonstration est achevée. ■

Ce résultat s'applique d'une façon naturelle à une suite de variables aléatoires fortement stationnaire.

Définition 3 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On dit qu'une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ est fortement stationnaire si $\forall m \in \mathbb{N}$, les processus $(X_n)_{n \in \mathbb{N}}$ et $(X_n)_{n \geq m}$ ont même loi.

On a alors le résultat suivant :

Corollaire 1 Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires fortement stationnaire vérifiant :

$$\mathbb{E}|X_1 + X_2 + \cdots + X_N|^p = O(\Psi(N))$$

où $\frac{\Psi(N)}{N}$ est une fonction non décroissante, alors pour tout $\epsilon > 0$, on a :

$$X_1(\omega) + X_2(\omega) + \cdots + X_N(\omega) = o\left((\Psi(N) (\log(N))^{p+1+\epsilon})^{\frac{1}{p}}\right) \quad \mathbb{P}(d\omega) - p.s.$$

Preuve : Puisque $(X_i)_{i \in \mathbb{N}}$ est fortement stationnaire alors $\forall M \in \mathbb{N}$:

$$\mathbb{E}|X_{M+1} + X_{M+2} + \cdots + X_{M+N}|^p = \mathbb{E}|X_1 + X_2 + \cdots + X_N|^p = O(\Psi(N)).$$

Le résultat se déduit du théorème précédent. ■

Remarque :

On utilise souvent ce théorème dans le cas $\Psi(N) = N$ et $p = 2$; ce qui conduit au résultat :

$$Si \quad \int_X S(M, N, x)^2 dx = O(N) \quad \text{uniformément en } M,$$

alors pour tout $\epsilon > 0$

$$S(0, N, x) = o\left(N^{\frac{1}{2}}(\log(N))^{\frac{3}{2}+\epsilon}\right) \quad \mu(dx) - p.s..$$

Pour une suite de variables aléatoires, on obtient le corollaire suivant.

Corollaire 2 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires appartenant à L^2 , vérifiant

$$\mathbb{E}|X_{M+1} + X_{M+2} + \cdots + X_{M+N}|^2 = O(N) \quad \text{uniformément en } M, \quad (3.8)$$

alors pour tout $\varepsilon > 0$

$$X_1(\omega) + X_2(\omega) + \cdots + X_N(\omega) = o\left(N^{\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}\right) \quad \mathbb{P}(d\omega) - p.s.$$

Preuve : Il suffit de prendre $\Psi(N) = N$ et $p = 2$. ■

Exemple : Les processus L^2 -stationnaires.

Considérons maintenant un système dynamique $(X, \mathcal{A}, \lambda, \tau)$. On peut déduire du théorème précédent un résultat ergodique fort qui nous suggère une estimation de la vitesse de convergence, dans le théorème ergodique ponctuel de Birkhoff.

Proposition 1 Soient $(X, \mathcal{A}, \lambda, \tau)$ un système dynamique et f une fonctionnelle de X , d'intégrale nulle et telle que :

$$\int_X \left| \sum_{n=0}^{N-1} f \circ \tau^n \right|^2 d\lambda = O(N) \quad (3.9)$$

alors

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} f \circ \tau^n(x) = o(N^{-\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}) \quad \lambda(dx) - p.s.. \quad (3.10)$$

Preuve : Cette proposition découle du corollaire précédent (corollaire 2). En effet, utilisons la propriété (3.9) et le fait que τ préserve la mesure λ pour vérifier la propriété (3.8). C'est à dire :

$$\int_X S(M, N, x)^2 d\lambda = \int_X \left| \sum_{n=0}^{N-1} f \circ \tau^n \right|^2 d\tau^M \lambda = \int_X \left| \sum_{n=0}^{N-1} f \circ \tau^n \right|^2 d\lambda = O(N).$$

Par suite, les hypothèses du corollaire 2 sont vérifiées. Donc pour tout $\varepsilon > 0$

$$\sum_{n=0}^{N-1} f \circ \tau^n = o(N^{\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}) \quad \lambda - p.s.,$$

soit encore

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ \tau^n = o(N^{-\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}) \quad \lambda - p.s..$$

■

Dans le cas où l'intégrale de la fonctionnelle f est non nulle, on utilise le corollaire précédent comme suit.

Corollaire 3 Soient $(X, \mathcal{A}, \lambda, \tau)$ un système dynamique et f une fonctionnelle de X , telle que

$$Var(\sum_{n=0}^{N-1} f \circ \tau^n) = O(N), \quad (3.11)$$

alors

$$\frac{1}{N} \sum_{n=0}^{N-1} f \circ \tau^n - \int_X f d\lambda = o(N^{-\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}) \quad \lambda - p.s.. \quad (3.12)$$

Preuve : Il suffit d'appliquer la proposition précédente à $f - \int_X f d\lambda$. ■

Dans la suite, on donnera une hypothèse de base pour le résultat ergodique fort dans le cas d'une transformation mélangeante.

3.2 Sur la vitesse de Convergence du shift.

On considère un système dynamique $(X, \mathcal{A}, \tau, \lambda)$, avec τ une transformation mélangeante. On suppose que $\int_X F d\lambda = 0$ et on pose

$$\sigma^2 := \sigma^2(F) = \|F\|_2^2 + 2 \sum_{k=1}^{\infty} \langle F \circ \tau^k, F \rangle \quad (3.13)$$

si la série est convergente et $+\infty$ sinon.

On désigne par $\langle \cdot, \cdot \rangle$ le produit scalaire sur $L^2(X)$ et $\|\cdot\|_2$ la norme sur $L^2(X)$:

$$\langle F, G \rangle = \int_X FG d\lambda \quad \|F\|_2 = \left(\int_X F^2 d\lambda \right)^{\frac{1}{2}}.$$

Si F est d'intégrale non nulle on prend :

$$\sigma^2 := \sigma^2(F - \mathbb{E}F) = Var(F) + 2 \sum_{k=1}^{\infty} Cov(F \circ \tau^k, F) \quad (3.14)$$

si la série est convergente et $+\infty$ sinon, où Var désigne la variance et Cov la covariance. On a alors le résultat suivant :

Proposition 2 Soit $F \in L^2(X, \mathcal{A}, \lambda)$ d'intégrale nulle et telle que $\sigma^2(F)$ fini alors :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_X \left| \sum_{n=0}^{N-1} F \circ \tau^n \right|^2 = \sigma^2 \quad (3.15)$$

et par conséquent

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} F \circ \tau^n = o\left(N^{-\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon}\right) \quad \lambda - p.s..$$

Preuve : Une fois que le premier résultat est établi, le deuxième s'obtient en appliquant la proposition 1 de la section 2, corollaire du théorème de Gál et Koksma sur les systèmes dynamiques.

$$\begin{aligned}
\int_X \left| \sum_{n=0}^{N-1} F \circ \tau^n \right|^2 d\lambda &= \sum_{1 \leq k, l \leq N} \langle F \circ \tau^k, F \circ \tau^l \rangle \\
&= N \|F\|_2^2 + 2 \sum_{l=1}^N \left(\sum_{k=1}^{N-l} \langle F \circ \tau^k, F \rangle \right) \\
&= N \sigma^2 - 2 \sum_{l=1}^N \left(\sum_{k=N-l+1}^N \langle F \circ \tau^k, F \rangle \right) \\
&\quad - 2 \sum_{l=1}^N \left(\sum_{k=N+1}^{\infty} \langle F \circ \tau^k, F \rangle \right) \\
&= N \sigma^2 - 2 \sum_{k=1}^N k \langle F \circ \tau^k, F \rangle - 2N \sum_{k=N+1}^{\infty} \langle F \circ \tau^k, F \rangle
\end{aligned}$$

donc

$$\frac{1}{N} \int_X \left| \sum_{n=0}^{N-1} F \circ \tau^n \right|^2 d\lambda = \sigma^2 - \frac{2}{N} \sum_{k=1}^N k \langle F \circ \tau^k, F \rangle - 2 \sum_{k=N+1}^{\infty} \langle F \circ \tau^k, F \rangle.$$

Le deuxième terme, à droite de l'égalité, converge vers zéro, d'après le lemme de Kronecker (cf. par exemple [35]). Le troisième terme, lui, converge vers zéro, comme reste d'une série convergente. On a donc

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_X \left| \sum_{n=0}^{N-1} F \circ \tau^n \right|^2 d\lambda = \sigma^2 \quad C.Q.F.D.$$

■

Remarque : On a utilisé seulement que la série $\sigma^2(F)$ est convergente et que τ préserve la mesure λ . Cependant si τ n'est pas mélangeante, la classe des fonctions F tel que $\sigma^2(F) < \infty$ ne contient aucune fonction de quelque intérêt.

Reprenons notre système dynamique $([0, 1]^{\mathbb{N}}, \mathcal{B}([0, 1]^{\otimes \mathbb{N}}), \lambda, \theta)$. Rappelons la définition de σ^2 dans le cas de la transformation du décalage.

$$\sigma^2 := \sigma^2(F) = \text{Var}(F) + 2 \sum_{k=1}^{\infty} \text{Cov}(F \circ \theta^k, F). \quad (3.16)$$

Si de plus F est d'intégrale nulle alors :

$$\sigma^2 := \sigma^2(F) = \|F\|_2^2 + 2 \sum_{k=1}^{\infty} \langle F \circ \theta^k, F \rangle \quad (3.17)$$

Il découle de la proposition précédente le résultat théorique principal sur la vitesse de convergence dans le théorème ergodique pour l'opérateur de décalage. Ce dernier sera considéré comme hypothèse de base pour déterminer des classes de fonctionnelles utiles pour les applications.

Proposition 3 Soit $F \in L^2([0, 1]^N, \mathcal{B}([0, 1]^N), \lambda, \theta)$, d'intégrale nulle et telle que $\sigma^2(F)$ est fini alors :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_{[0, 1]^N} \left| \sum_{n=0}^{N-1} F \circ \theta^n \right|^2 = \sigma^2 \quad (3.18)$$

et par conséquent

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n = o(N^{-\frac{1}{2}} (\log(N))^{\frac{3}{2} + \varepsilon}) \quad \lambda - p.s..$$

Preuve : Ce résultat est un simple corollaire de la proposition précédente. En effet, il suffit de remarquer que la transformation du décalage est une transformation mélangeante. Ceci a été fait dans le premier chapitre. ■

Remarques :

1/Les fonctions dépendant d'un nombre fini de variables, dites fonctions cylindriques, vérifient la condition $\sigma^2 < \infty$. En effet, si F ne dépend que des N premières coordonnées et si l'on suppose que $\int_{[0, 1]^N} F d\lambda = 0$, alors :

$$\sigma^2(F) = \|F\|_2^2 + 2 \sum_{k=1}^{N-1} \langle F \circ \theta^k, F \rangle.$$

2/On sait que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires réelles indépendantes équidistribuées d'intégrale nulle et telle que $\mathbb{E}X_1^2 < \infty$, alors

$$\frac{1}{N} (X_1 + \dots + X_N) = O(N^{-\frac{1}{2}} (\log \log(N))^{\frac{1}{2}}),$$

L'hypothèse d'indépendance a été affaiblie dans plusieurs travaux (cf. Berger [7]).

3/Il est évident que les estimations de la remarque précédente, connues sous le nom de propriété du logarithme itéré, sont plus fortes que le résultat de la proposition 3. L'intérêt de notre résultat est de donner une estimation assez voisine du logarithme itéré mais sous des hypothèses plus faibles et assez naturelles dans le cadre de la simulation.

En simulation on rencontre essentiellement deux types de fonctionnelles dépendant d'un nombre infini de coordonnées :

- Soit la fonctionnelle F ne dépend que d'un nombre fini de coordonnées, le long de chaque trajectoire $U(\omega) = (U_1(\omega), U_2(\omega), \dots, U_k(\omega), \dots)$; cette classe fera l'objet du troisième paragraphe.
- Soit l'on arrive à approcher F à une vitesse suffisante par une suite de fonctions $(F_k)_{k \in \mathbb{N}}$ ne dépendant que d'un nombre fini de coordonnées; ceci fait l'objet du théorème suivant.

Théorème 4 Soit F une fonctionnelle d'intégrale nulle et ayant un moment d'ordre 2. Si l'on suppose qu'il existe $(F_k)_{k \in \mathbb{N}}$ une suite de $L^2([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$, telle que F_k est $\sigma(U_1, \dots, U_k)$ -mesurable et que F_k converge vers F dans $L^2([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$ avec $\sum_{k=1}^{\infty} \|F - F_k\|_2 < \infty$ alors

$$| \langle F \circ \theta^k, F \rangle | \leq \|F\|_2 \|F - F_k\|_2. \quad (3.19)$$

Par conséquent $\sigma^2(F)$ est une série absolument convergente et

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n = o(N^{-\frac{1}{2}} (\log(N))^{\frac{3}{2} + \varepsilon}) \quad \lambda - p.s.$$

Preuve : L'estimation de la vitesse de convergence découle de la proposition 3. Pour ce faire on démontre que la série, définissant $\sigma^2(F)$ dans (3.17), est absolument convergente. En effet

$$\langle F \circ \theta^k, F \rangle = \langle F \circ \theta^k, F_k \rangle + \langle F \circ \theta^k, F - F_k \rangle.$$

En remarquant que $F \circ \theta^k$ est $\sigma(U_{k+1}, \dots)$ -mesurable et que F_k est $\sigma(U_1, \dots, U_k)$ -mesurable, on a $F \circ \theta^k$ et F_k sont indépendantes, ainsi

$$\langle F \circ \theta^k, F_k \rangle = \int_{[0,1]^N} F \circ \theta^k d\lambda \int_{[0,1]^N} F_k d\lambda = 0,$$

et par suite, d'après l'inégalité de Cauchy-Schwartz et l'invariance de λ par θ :

$$\begin{aligned} | \langle F \circ \theta^k, F \rangle | &= | \langle F \circ \theta^k, F - F_k \rangle | \\ &\leq \|F\|_2 \|F - F_k\|_2. \end{aligned}$$

D'où le résultat annoncé. ■

Remarque :

Il est équivalent de prendre $F_k = \mathbb{E}(F/\sigma(U_1, \dots, U_k))$, puisque l'espérance conditionnelle minimise $\|F - G\|_2$ pour toutes $G \in L^2$.

Exemples :

1/ Soit f une fonction de carré intégrable sur $[0, 1]$ et d'intégrale nulle. Si l'on pose

$$F = \sum_{k=1}^{\infty} a_k f(U_k) \quad \text{avec} \quad \sum_{i=1}^{\infty} a_i^2 < \infty,$$

et si l'on considère la suite de variables aléatoires,

$$F_k = \sum_{i=1}^k a_i f(U_i),$$

alors la condition $\sum_{k=1}^{\infty} \|F - F_k\|_2 < \infty$, devient

$$\sum_{l=1}^{\infty} \left(\sum_{i>l} a_i^2 \right)^{\frac{1}{2}} < \infty.$$

En particulier, pour $a_n = \frac{1}{n^\alpha}$ cette série converge si et seulement si $\alpha > \frac{3}{2}$.

2/ On dira qu'une fonction F est lipischitzienne sur $[0, 1]^N$, s'il existe une constante $M \in \mathbb{R}$ tel que $|F(U) - F(U')| \leq M d(U, U') \quad \forall U, U' \in [0, 1]^N$ avec $d(U, U') = \sum_{k=1}^{\infty} \frac{|U_k - U'_k|}{2^k}$.

Soit F une fonction lipischitzienne sur $[0, 1]^N$, si l'on considère la fonction F_k de $[0, 1]^k$ dans \mathbb{R} définie par

$$F_k(x_1, \dots, x_k) = \int_{[0,1]^N} F(x_1, \dots, x_k, U_1, U_2, \dots) \lambda(dU)$$

et si l'on note par Π_k la projection canonique de $[0, 1]^N$ dans $[0, 1]^k$, alors F est approchée dans L^2 par la suite de fonctions $(F_k \circ \Pi_k)_{k \in \mathbb{N}}$. De plus, on a :

$$\begin{aligned} |F(x) - F_k \circ \Pi_k(x)| &\leq \int_{[0,1]^N} |F(x) - F(x_1, x_2, \dots, x_k, U_1, U_2, \dots)| \lambda(dU) \\ &\leq M \int_{[0,1]^N} \sum_{n=k+1}^{\infty} \frac{|U_{n-k} - x_n|}{2^n} \lambda(dU) \\ &\leq \frac{M}{2^k} \sum_{k=1}^{\infty} \frac{1}{2^{n-1}} = \frac{M}{2^{k-1}}. \end{aligned}$$

La condition $\sum_{k=1}^{\infty} \|F - F_k \circ \Pi_k\|_2 < \infty$ est donc vérifiée. Il en résulte que, pour toute fonction lipischitzienne sur $[0, 1]^N$ on a :

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n = o(N^{-\frac{1}{2}} (\log(N))^{\frac{3}{2} + \varepsilon}) \quad \lambda - p.s.$$

3/ Soit F une fonctionnelle sur $[0, 1]^N$, \mathcal{F}_T -mesurable, avec T un $(\sigma(U_1, \dots, U_n))_{n \in \mathbb{N}}$ -temps d'arrêt. Si l'on suppose que F est d'intégrale nulle et de carré intégrable et si l'on considère la suite de variables aléatoires $F_k = F \mathbf{1}_{\{T \leq k\}}$ alors la condition, $\sum_{k=1}^{\infty} \|F - F_k\|_2 < \infty$, devient

$$\sum_{k=1}^{\infty} \|F \mathbf{1}_{\{T > k\}}\|_2 < \infty. \quad (3.20)$$

En appliquant l'inégalité de Cauchy-Schwartz, on montre que la série converge pour les fonctions vérifiant : F admet un moment d'ordre $2p$ et T un moment d'ordre supérieur à q , avec $\frac{1}{p} + \frac{1}{q} = 1$.

Le principe du dispositif informatique et les exemples numériques, donnés dans la deuxième section du deuxième chapitre, montrent que la méthode du décalage a été conçue, essentiellement, pour le calcul d'espérance des fonctions \mathcal{F}_T -mesurable où T est un temps d'arrêt. Cette classe, étant la plus utile application, on consacrera le paragraphe suivant à l'étude de ces fonctions. Nous y améliorerons notablement le premier résultat (exemple 3) que nous venons d'obtenir.

3.3 Vitesse de convergence des fonctions \mathcal{F}_T mesurables.

Dans la grande majorité des cas les fonctionnelles sont issues de modèle non anticipatifs et se trouvent être \mathcal{F}_T -mesurables pour un temps aléatoire T qui est un temps d'arrêt. Or, dès que le temps d'arrêt admet un moment d'ordre plus grand que 2 les fonctionnelles \mathcal{F}_T -mesurables ne contiennent pas de cas pathologiques de type "convergence ultra-lente" comme ceux évoqués au chapitre précédent (section 2.3). Pour plus de généralité, on considérera, un temps d'arrêt T de $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et une fonctionnelle F , \mathcal{F}_T -mesurable, avec $\mathcal{F}_n = \sigma(U_1, \dots, U_n)$. On a alors le premier résultat suivant :

Théorème 5 *Si T admet un moment d'ordre $p > 2$ alors pour toute $F \in L^2([0, 1]^{\mathbb{N}}, \mathcal{F}_T)$, telle que $\int_{[0, 1]^{\mathbb{N}}} F = 0$, on a*

$$| \langle F \circ \theta^k, F \rangle | \leq \frac{\int_{[0, 1]^{\mathbb{N}}} |F|^2 d\lambda (\mathbb{E} T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}}, \quad (3.21)$$

par conséquent $\sigma^2(F)$ est une série absolument convergente et

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n = o(N^{-\frac{1}{2}} (\log(N))^{\frac{3}{2} + \varepsilon}) \quad \lambda - p.s.$$

Preuve : Il suffit d'établir l'équation (3.21), le reste est immédiat.

$$\langle F \circ \theta^k, F \rangle = \langle F \circ \theta^k, F \cdot 1_{\{T \leq k\}} \rangle + \langle F \circ \theta^k, F \cdot 1_{\{T > k\}} \rangle.$$

Comme dans le théorème 4, en remarquant que $F \circ \theta^k$ est $\sigma(U_{k+1}, \dots)$ -mesurable et que $F \cdot 1_{\{T \leq k\}}$ est $\sigma(U_1, \dots, U_k)$ -mesurable, on déduit que $F \circ \theta^k$ et $F \cdot 1_{\{T \leq k\}}$ sont indépendantes. D'où

$$\langle F \circ \theta^k, F \cdot 1_{\{T \leq k\}} \rangle = \int_{[0, 1]^{\mathbb{N}}} F \circ \theta^k d\lambda \int_{[0, 1]^{\mathbb{N}}} F \cdot 1_{\{T \leq k\}} d\lambda = 0,$$

et par suite

$$\begin{aligned} | \langle F \circ \theta^k, F \rangle | &= | \langle F \circ \theta^k, F \cdot 1_{\{T > k\}} \rangle | \\ &= | \langle F \circ \theta^k \cdot 1_{\{T > k\}}, F \cdot 1_{\{T > k\}} \rangle | \\ &\leq \left(\int_{[0, 1]^{\mathbb{N}}} |F \circ \theta^k \cdot 1_{\{T > k\}}|^2 d\lambda \right)^{\frac{1}{2}} \cdot \left(\int_{[0, 1]^{\mathbb{N}}} |F \cdot 1_{\{T > k\}}|^2 d\lambda \right)^{\frac{1}{2}}. \end{aligned} \quad (3.22)$$

On obtient là une majoration plus fine que celle du théorème 4. Ceci est dû au fait que F est \mathcal{F}_T -mesurable et donc les fonctions F_k sont égales à $F \cdot 1_{\{T \leq k\}}$. A ce stade on constate que

$F \circ \theta^k$ est $\sigma(U_{k+1}, \dots)$ -mesurable, et que $\{T > k\}$ appartient à $\sigma(U_1, \dots, U_k)$, il apparaît que $F \circ \theta^k$ et $1_{\{T > k\}}$ sont indépendantes. Par suite

$$\begin{aligned} \int_{[0,1]^N} |F \circ \theta^k \cdot 1_{\{T > k\}}|^2 d\lambda &= \int_{[0,1]^N} |F \circ \theta^k|^2 d\lambda \cdot \int_{[0,1]^N} 1_{\{T > k\}} d\lambda \\ &= \int_{[0,1]^N} |F|^2 d\lambda \cdot \mathbb{P}(T > k), \end{aligned}$$

d'où

$$\begin{aligned} | \langle F \circ \theta^k, F \rangle | &\leq \left(\int_{[0,1]^N} |F|^2 d\lambda \right)^{\frac{1}{2}} \cdot (\mathbb{P}(T > k))^{\frac{1}{2}} \cdot \left(\int_{[0,1]^N} |F \cdot 1_{\{T > k\}}|^2 d\lambda \right)^{\frac{1}{2}} \\ &\leq \int_{[0,1]^N} |F|^2 d\lambda \cdot (\mathbb{P}(T > k))^{\frac{1}{2}}. \end{aligned}$$

Comme $T \in L^p$, il vient alors

$$| \langle F \circ \theta^k, F \rangle | \leq \frac{C}{k^{\frac{p}{2}}}$$

avec

$$C = \left(\int_{[0,1]^N} |F|^2 d\lambda \right) (\mathbb{E}T^p)^{\frac{1}{2}},$$

Ce qui achève la démonstration. ■

Remarques :

1/ Les techniques utilisées à la fin de la démonstration exploitent les propriétés d'une fonction \mathcal{F}_T -mesurable. Ainsi, le résultat obtenu n'entre pas dans le cadre du théorème 4.

2/D'après la démonstration, on observe que la condition

$$\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{2}} \cdot \left(\int_{[0,1]^N} |F \cdot 1_{\{T > k\}}|^2 d\lambda \right)^{\frac{1}{2}} < \infty, \quad (3.23)$$

est en fait suffisante pour avoir le résultat.

3/Contrairement aux apparences majorer, comme on le fait dans la démonstration, le terme $\int_{[0,1]^N} |F \cdot 1_{\{T > k\}}|^2 d\lambda$ par $\int_{[0,1]^N} |F|^2 d\lambda$ n'est pas grave car ce que l'on perd est récupéré dans l'énoncé un peu plus général du théorème ci-dessous.

4/Sous les hypothèses du théorème 5 on a :

$$\begin{aligned} \sigma^2(F) &\leq \|F\|_2^2 \left(1 + 2\sqrt{\mathbb{E}T^p} \sum_{k=1}^{\infty} \frac{1}{k^{\frac{p}{2}}} \right) \\ &\leq \|F\|_2^2 \left(1 + 2\sqrt{\mathbb{E}T^p} \int_1^{\infty} \frac{dx}{x^{\frac{p}{2}}} \right) \end{aligned}$$

On a donc dans ce cas un majorant relativement explicite pour $\sigma^2(F)$ donné par :

$$\sigma^2(F) \leq \|F\|_2^2 \left(1 + \frac{4}{p-2} \sqrt{\mathbb{E}T^p} \right) \quad (3.24)$$

Souvent, dans la pratique F admet des moments à tout ordre, par exemple : si l'on prend la marche aléatoire réelle de loi normale et T le temps de sortie d'un intervalle et si l'on considère $F = T$ ou $F = X_T$ alors F admet des moments d'ordre $p \quad \forall p \in \mathbb{N}$. Cependant, il apparaît, d'après la démonstration du théorème 5, que si la fonctionnelle F admet un moment supérieure à 2 alors on a une condition plus faible sur le moment du temps d'arrêt T . On obtient le résultat suivant :

Théorème 6 *Pour toute $F \in L^q([0,1]^N, \mathcal{F}_T)$, $q \geq 2$, telle que $\int_{[0,1]^N} F d\lambda = 0$, si T vérifie $\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{p}} < \infty$ avec $\frac{1}{p} + \frac{1}{q} = 1$, on a :*

$$| \langle F \circ \theta^k, F \rangle | \leq \left(\int_{[0,1]^N} |F|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}} \quad (3.25)$$

par conséquent $\sigma^2(F)$ est une série absolument convergente et

$$\forall \varepsilon > 0 \quad \frac{1}{N} \sum_{n=0}^{N-1} F \circ \theta^n = o(N^{-\frac{1}{2}} (\log(N))^{\frac{3}{2}+\varepsilon}) \quad \lambda - p.s.$$

Preuve : On remarque que $q \geq 2$ et par suite $p \leq q$. En reprenant donc la démonstration du théorème 5 et en appliquant l'inégalité de Hölder au niveau de l'équation(3.22) en lieu et place de l'inégalité de Cauchy-Schwartz, on obtient le résultat suivant

$$| \langle F \circ \theta^k, F \rangle | \leq \left(\int_{[0,1]^N} |F \circ \theta^k \cdot 1_{\{T > k\}}|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F \cdot 1_{\{T > k\}}|^q d\lambda \right)^{\frac{1}{q}}.$$

On conclut comme dans le théorème précédent et on obtient :

$$| \langle F \circ \theta^k, F \rangle | \leq \left(\int_{[0,1]^N} |F|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}}$$

$F \in L^p$ car on a pris $F \in L^q$ et $p \leq q$ puisque $q \geq 2$. Ceci achève la démonstration. ■

Remarques :

- Si $p > 1$, $\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{p}} < \infty$ par exemple dès que $T \in L^{p+\varepsilon}([0,1]^N)$ pour $\varepsilon > 0$.
- Si $p > 1$, $T \in L^{p+\varepsilon}([0,1]^N)$ pour $\varepsilon > 0$, $F \in L^q([0,1]^N)$ d'intégrale nulle et $\frac{1}{p} + \frac{1}{q} = 1$ on a un majorant relativement explicite pour $\sigma^2(F)$:

$$\begin{aligned} \sigma^2(F) &\leq \|F\|_2^2 + 2\|F\|_p \|F\|_q \frac{(\mathbb{E}T^{p+\varepsilon})^{\frac{1}{p}}}{k^{1+\frac{\varepsilon}{p}}} \\ \sigma^2(F) &\leq \|F\|_2^2 + \frac{2p}{\varepsilon} \|F\|_p \|F\|_q (\mathbb{E}T^{p+\varepsilon})^{\frac{1}{p}} \end{aligned} \quad (3.26)$$

- Si $p = 1$, $\sum_{k=1}^{\infty} \mathbb{P}(T > k) = \mathbb{E}T < \infty$ si et seulement si $T \in L^1([0, 1]^N)$.
- Si $p = 1$ et $F \in L^\infty([0, 1]^N)$ d'intégrale nulle on a :

$$\begin{aligned} \sigma^2(F) &\leq \|F\|_2^2 + 2 \int_{[0,1]^N} |F(U)| d\lambda \sup_{U \in [0,1]^N} |F(U)| \mathbb{E}T \\ &\leq \|F\|_\infty^2 (1 + 2\mathbb{E}T) \end{aligned} \quad (3.27)$$

On a donc dans ce cas un majorant relativement explicite pour $\sigma^2(F)$.

Un exemple courant de fonctionnelle bornée qui rentre dans le cadre de nos applications est de prendre $F = X_T$, où T est le temps d'entrée dans un compact.

3.4 Illustration informatique.

Nous avons voulu vérifier la validité du résultat obtenu dans le paragraphe précédent et tester l'existence éventuelle d'une loi du logarithme itéré "classique" en faisant des simulations. Nous avons considéré des marches aléatoires définies de la façon suivante :

$$X_0 = x, \quad X_{n+1} = X_n + h(U_{n+1}^1, \dots, U_{n+1}^d),$$

où les $(U_n^i)_{n,i \in \mathbb{N}}$ sont des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Pour chaque marche aléatoire, on considère le temps d'arrêt $T = \inf\{n \geq 1; X_n \notin [a, b]\}$ et des fonctionnelles de type $F = G(X_T, T)$.

En dehors des majorations que l'on peut éventuellement obtenir à partir des remarques faites à la page 35 et page 37, l'estimation de σ s'obtient à partir de l'équation 3.18, démontrée dans la proposition 3, à savoir que :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Var} \left(\sum_{n=0}^{N-1} F \circ \theta^n \right) = \sigma^2.$$

L'utilisation du lemme de Kronecker dans la preuve de ce résultat ne favorise pas une détermination de l'erreur théorique. Ainsi nous n'avons pas actuellement de méthode plus fiable que d'attendre une stabilisation numérique de cette suite et d'assimiler la valeur stable à sa limite. Si l'on note par N_{\max} la valeur à partir de laquelle les premiers chiffres des termes de la suite ne sont plus modifiés, on prendra comme estimation de σ la quantité, $\hat{\sigma} = \frac{1}{N_{\max}} \text{Var} \left(\sum_{n=0}^{N_{\max}-1} F \circ \theta^n \right)$. Pour N fixé, on simule $\text{Var} \left(\sum_{n=0}^{N-1} F \circ \theta^n \right)$ par la méthode de Monte Carlo sur 5000 trajectoires indépendantes. Une fois $\hat{\sigma}$ assimilé à σ , une simulation des trajectoires du processus

$$B_n(F) = \frac{F + F \circ \theta + \dots + F \circ \theta^{n-1} - n\mathbb{E}F}{\hat{\sigma} \sqrt{n(\log n)^3}},$$

doit alors confirmer sa convergence vers 0. De toute façon, une erreur sur σ n'empêche pas le principe de la vérification de la vitesse de convergence puisqu'elle ne modifie le rapport que par un facteur constant.

Exemples : Considérons la marche aléatoire :

$$X_0 = 0, X_{n+1} = X_n + \frac{2U_{n+1} - 1}{2} \text{ et } T = \inf\{n \geq 1, X_n \notin [-5, 1]\}.$$

- Pour $F = X_T$:

En prenant $N_{max} = 10\,000$, on obtient $\hat{\sigma} = 23.724$. Le tableau suivant représente la convergence vers 0 de la suite $B_n(X_T)$.

n	1 000	10 000	100 000	200 000
$B_n(X_T)$	0.0249	-0.0032	-0.0358	-0.0231
n	400 000	600 000	800 000	1 000 000
$B_n(X_T)$	-0.0238	-0.0014	0.0050	-0.0031

Tableau 3.1

- Pour $F = T$:

En prenant $N_{max} = 10\,000$, on obtient $\hat{\sigma} = 574.281$. Le tableau suivant représente la convergence vers 0 de la suite $B_n(T)$.

n	1 000	10 000	100 000	200 000
$B_n(T)$	-0.0242	0.0180	0.0383	0.0236
n	400 000	600 000	800 000	1 000 000
$B_n(T)$	0.0180	0.0067	-0.0037	-0.0057

Tableau 3.2

La représentation graphique ci-dessous illustre la même simulation.

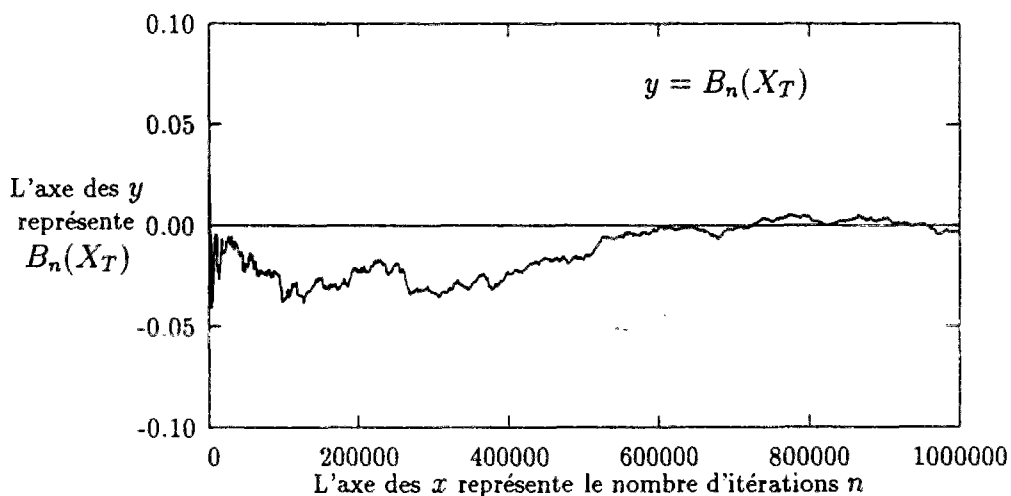


Figure 3.1

La représentation graphique ci-dessous illustre la même simulation.

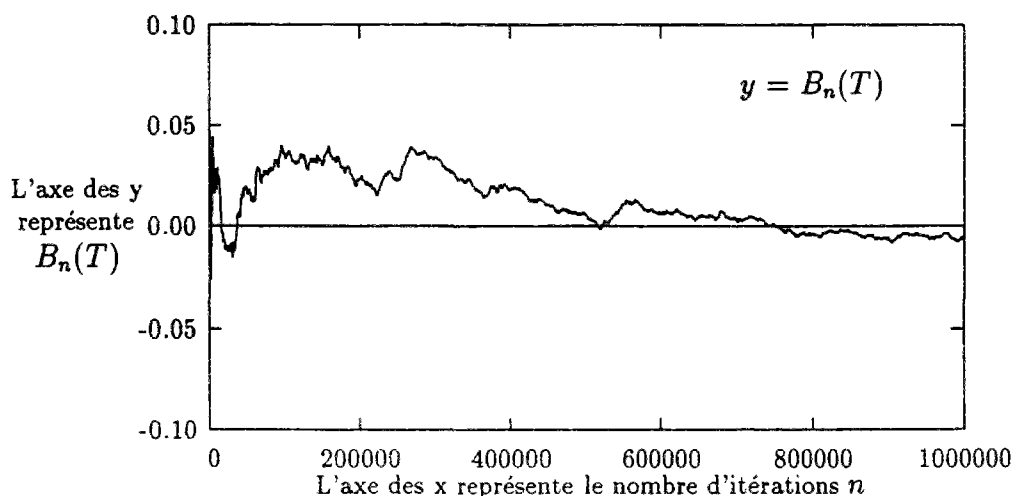


Figure 3.2

Il paraît naturel d'essayer de vérifier expérimentalement si la loi du logarithme itéré est satisfaite. Cette loi a fait l'objet de nombreux travaux, voir Berger [7] dans le cas d'un espace de Banach ; Walter Phillip et William Stout dans [42] ont étudié la somme partielle de plusieurs suites de variables aléatoires faiblement dépendantes et ils ont proposés des principes d'invariances. Nous verrons au chapitre 5 que la loi du logarithme itéré s'applique sous des hypothèses plus restrictives sur les moments de F et T . Pour ce faire on a simulé des trajectoires du processus :

$$C_n(F) = \frac{F + F \circ \theta + \dots + F \circ \theta^{n-1} - n\mathbb{E}F}{\hat{\sigma}\sqrt{n \log \log n}}.$$

On remarque clairement sur les simulations que les courbes ont changé d'allure : la plupart oscillent entre -1 et 1, sans se rapprocher d'une valeur particulière. Pour les exemples étudiés précédemment, les résultats obtenus sont représentés par la figure 3.3 et 3.4. Toutes les suites ont été simulées jusqu'à l'indice 10^6 . Ce nombre se trouve être assez grand pour observer une convergence par la méthode des tirages indépendants. Le fait que toutes les trajectoires restent entre -1 et 1, sans se rapprocher d'une valeur particulière pour 10^5 itérations, peut être considéré comme la meilleure observation numérique de la loi du logarithme itéré. En effet, il est difficile d'observer numériquement une limite supérieure ou limite inférieure. En plus, les nombres pseudo-aléatoires eux-mêmes ne vérifient pas la loi du logarithme itéré. Ces résultats positifs pour les différents exemples traités nous suggèrent que la loi du logarithme itéré est susceptible d'être satisfaite pour une sous-classe de fonctions vérifiant $\sigma^2 < \infty$. Nous y reviendrons au cinquième chapitre.

- Pour $F = X_T$:

Les réalisations du processus, obtenues par simulation, ont cette allure :

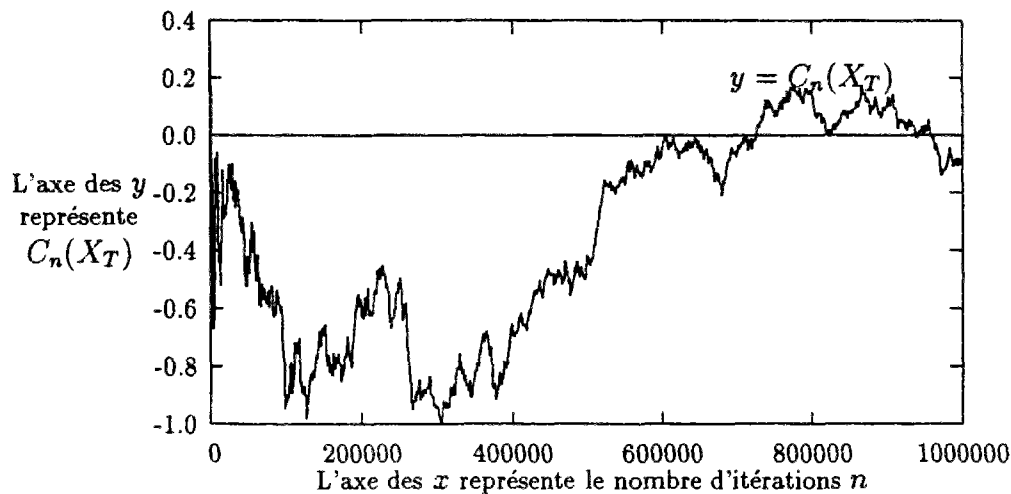


Figure 3.3

- Pour $F = T$:

Les réalisations du processus, obtenues par simulation, ont cette allure :

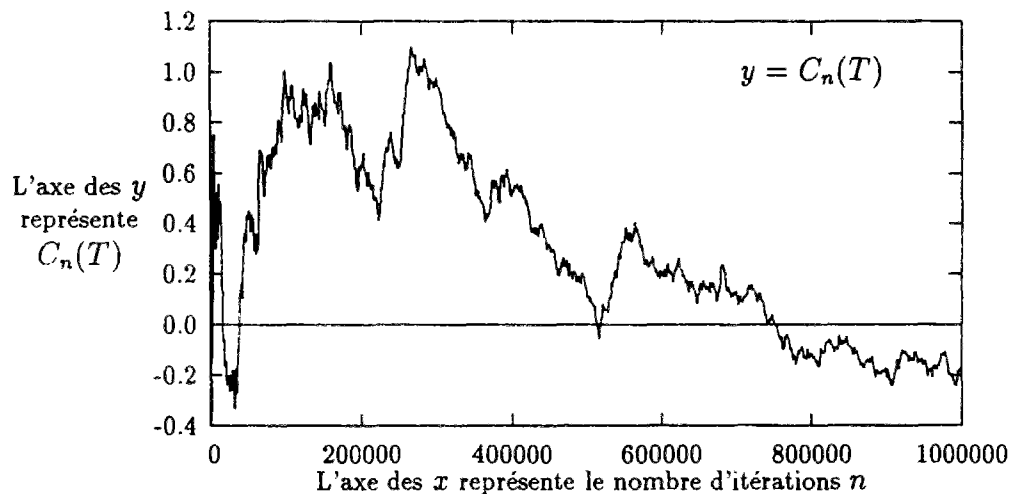


Figure 3.4

3.5 Conclusion.

Les résultats établis ci-dessus montrent que la méthode du décalage pour le calcul de l'espérance de fonctionnelles de processus aléatoires a une vitesse asymptotique mathématique du même ordre de grandeur en pratique que la loi des grands nombres, pour une large classe de fonctionnelles parmi les plus couramment rencontrées. En effet, l'erreur théorique de la

méthode du shift et celle de Monte Carlo “tournent autour” $O(\frac{1}{\sqrt{N}})$. Ceci renforce l’intérêt de cette méthode dont l’avantage essentiel, il convient de le souligner, réside dans l’économie de calculs et de nombres aléatoires que permet son implémentation informatique et permet, donc, une vitesse informatique plus grande que la méthode de Monte Carlo.

Chapitre 4

Convergence en loi de l'erreur normalisée.

Le but de ce chapitre est d'étudier le comportement en loi de l'erreur de cette méthode d'intégration, plus précisément on déterminera la loi asymptotique de $\frac{1}{\sqrt{N}}(\sum_{n=0}^{N-1} F \circ \theta^n - N \mathbb{E}F)$ pour certaines classes de fonctions. Dans le premier paragraphe on s'intéressera aux fonctionnelles F couramment employées en pratique. On montre notamment que si T admet un moment d'ordre $p > 2$ et si F est une fonction de L^2 , \mathcal{F}_T mesurable et d'intégrale nulle, alors on a le théorème de la limite centrale à savoir que :

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1), \quad (4.1)$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

Dans la littérature on démontre le théorème de la limite centrale pour des processus dépendants mais sous des hypothèses moins bien adaptées aux situations pratiques rencontrées en simulation, par exemple les processus dits φ -mélangeants (voir Billingsley [9]). Dans le deuxième paragraphe on montrera par des techniques différentes que les fonctions lipschitziennes vérifient aussi le théorème de la limite centrale. Ces techniques ont été utilisées pour une classe de transformations définies sur l'intervalle $[0, 1]$ (voir [48]). Les fonctions lipschitziennes sont moins utilisées en pratique, cependant, la démonstration nous paraît intéressante et permet d'obtenir une vitesse de convergence dans le théorème de la limite centrale de type Berry-Essen. Le troisième paragraphe illustre par des simulations numériques les résultats théoriques obtenus.

4.1 Théorème de la limite centrale pour les fonctions dépendant d'un temps d'arrêt.

Tout d'abord nous allons rappeler quelques résultats sur les processus φ -mélangeants (voir Billingsley [9]). Les techniques utilisées nous ont été utiles pour démontrer nos théorèmes. Cependant, notre suite $(F \circ \theta^k)_{k \in \mathbb{N}}$ ne rentre pas dans le cadre de ces processus. Nous allons

donner la définition d'un processus φ -mélangeant. Soit

$$\cdots, X_{-1}, X_0, X_1, \cdots \quad (4.2)$$

une suite de variables aléatoires strictement stationnaire sur un espace de probabilité quelconque $(\Omega, \mathcal{A}, \mathbb{P})$. Si l'on note par \mathcal{M}_a^b la tribu engendrée par les $(X_i)_{a \leq i \leq b}$, \mathcal{M}_a^∞ la tribu engendrée par les $(X_i)_{i \geq a}$ et \mathcal{M}_∞^b la tribu engendrée par les $(X_i)_{i \leq b}$ et si l'on considère une fonction φ sur \mathbb{N} à valeurs positives, vérifiant $\lim_{n \rightarrow \infty} \varphi(n) = 0$, alors $(X_i)_{i \in \mathbb{N}}$ est dit φ -mélangeant si pour tous k et $n \in \mathbb{N}$ avec $n \geq 1$:

$$\sup \left\{ |\mathbb{P}(E_2 | E_1) - \mathbb{P}(E_2)| : E_1 \in \mathcal{M}_\infty^k, E_2 \in \mathcal{M}_{k+n}^\infty \right\} \leq \varphi(n). \quad (4.3)$$

Soit encore $\forall k, n \in \mathbb{N}, n \geq 1, E_1 \in \mathcal{M}_\infty^k$ et $E_2 \in \mathcal{M}_{k+n}^\infty$ on a :

$$|\mathbb{P}(E_2 \cap E_1) - \mathbb{P}(E_2)\mathbb{P}(E_1)| \leq \varphi(n) \mathbb{P}(E_1).$$

Si $\varphi(n)$ est petit alors E_2 est pratiquement indépendant de E_1 . Dans un processus φ -mélangeant le futur lointain est pratiquement indépendant du présent et du passé. Billingsley a démontré le théorème de la limite centrale fonctionnelle et donc le théorème de la limite centrale pour des processus φ -mélangeant (voir [9]). Par suite on a le résultat suivant.

Théorème 7 *On suppose que $(X_i)_{i \in \mathbb{N}}$ est φ -mélangeant avec $\sum_{n=1}^{\infty} \varphi(n)^{\frac{1}{2}} < \infty$ et que X_0 est d'intégrale nulle et de variance finie. Alors la série*

$$\sigma^2 = \mathbb{E}(X_0^2) + 2 \sum_{k=1}^{\infty} \mathbb{E}(X_0 X_k)$$

est absolument convergente. Si de plus $\sigma^2 > 0$ alors

$$\frac{X_0 + X_1 + \cdots + X_{n-1}}{\sigma \sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1),$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

Preuve : Voir Billingsley [9]. ■

Exemple :

Une suite de variables aléatoires est dite m -dépendantes si les vecteurs (X_i, \dots, X_k) et $(X_{i+n}, \dots, X_{k+n})$ sont indépendants pour tous $n > m$. Dans cette terminologie, un processus indépendant est un processus 0-dépendant. Un processus m -dépendant est un processus φ -mélangeant avec $\varphi(n) = 0 \quad \forall n > m$ et il vérifie donc le théorème de la limite centrale.

Reprenons notre système dynamique $([0, 1]^{\mathbb{N}}, \mathcal{B}([0, 1]^{\mathbb{N}}), \lambda, \theta)$. Le processus $(F \circ \theta^k)_{k \in \mathbb{N}}$ associé à une fonction F ne dépendant que des N premières coordonnées est un processus N -dépendant. Par conséquent, les fonctions cylindriques vérifient le théorème de la limite centrale.

Proposition 4 Soit F une fonction cylindrique ne dépendant que des N premières coordonnées d'intégrale nulle et de variance finie. On a $\sigma^2(F) = \|F\|_2^2 + 2 \sum_{k=1}^{N-1} \langle F \circ \theta^k, F \rangle$ et si $\sigma^2 > 0$ alors :

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1), \quad (4.4)$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite.

Preuve : Le processus $(F \circ \theta^k)_{k \in \mathbb{N}}$ est un processus φ -mélangeant avec $\varphi(n) = 0 \quad \forall n > N$. Il vérifie donc les hypothèses du théorème précédent. ■

Etudions maintenant les classes de fonctions introduites dans le troisième chapitre. Commençons par les fonctions les plus utiles pour les applications. On a un temps d'arrêt T de $(\mathcal{F}_n)_{n \in \mathbb{N}}$ et une fonctionnelle F , \mathcal{F}_T -mesurable, avec $\mathcal{F}_n = \sigma(U_1, \dots, U_n)$. Traitons d'abord le cas où F admet un moment d'ordre 2.

Théorème 8 Si T admet un moment d'ordre $p > 2$ alors pour toute $F \in L^2([0, 1]^N, \mathcal{F}_T)$, telle que $\int_{[0, 1]^N} F = 0$ et $\sigma^2(F) > 0$ on a

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1), \quad (4.5)$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite.

Pour établir le théorème de la limite centrale, on calculera les limites de $\sigma^2(F.1_{T \leq l})$ et $\sigma^2(F.1_{T > l})$ lorsque l tend vers l'infini. En effet, si pour $l \in \mathbb{N}$ on note par :

$$\sigma_l^2 := \sigma^2(F.1_{T \leq l}) = \text{Var}(F.1_{T \leq l}) + 2 \sum_{k=1}^{\infty} \text{Cov}(F.1_{T \leq l} \circ \theta^k, F.1_{T \leq l})$$

et

$$\tau_l^2 := \sigma^2(F.1_{T > l}) = \text{Var}(F.1_{T > l}) + 2 \sum_{k=1}^{\infty} \text{Cov}(F.1_{T > l} \circ \theta^k, F.1_{T > l}),$$

on a les résultats suivants.

Lemme 3 Si T admet un moment d'ordre $p > 2$, alors pour toute $F \in L^2([0, 1]^N, \mathcal{F}_T)$, telle que $\int_{[0, 1]^N} F = 0$, on a

$$\lim_{l \rightarrow \infty} \sigma_l^2 = \sigma^2. \quad (4.6)$$

Preuve : En effet

$$\sigma_l^2 = \sigma^2(F.1_{T \leq l}) = \text{Var}(F.1_{T \leq l}) + 2 \sum_{k=1}^{\infty} \text{Cov}(F.1_{T \leq l} \circ \theta^k, F.1_{T \leq l}).$$

D'après le théorème 5, il résulte que :

$$\begin{aligned} |Cov(F.1_{T \leq l} \circ \theta^k, F.1_{T \leq l})| &\leq \int_{[0,1]^N} |F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}|^2 d\lambda \frac{(\mathbb{E}T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}} \\ &\leq \int_{[0,1]^N} |F.1_{T \leq l}|^2 d\lambda \frac{(\mathbb{E}T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}} \\ &\leq \int_{[0,1]^N} |F|^2 d\lambda \frac{(\mathbb{E}T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}}. \end{aligned}$$

D'où σ_l^2 est une série absolument convergente uniformément en l et puisque chaque terme de la série converge vers $\langle F \circ \theta^k, F \rangle$, $\lim_{l \rightarrow \infty} \sigma_l^2 = \sigma^2$. ■

Lemme 4 : Si T admet un moment d'ordre $p > 2$ alors pour toute $F \in L^2([0,1]^N, \mathcal{F}_T)$, telle que $\int_{[0,1]^N} F = 0$, on a

$$\lim_{l \rightarrow \infty} \tau_l^2 = 0. \quad (4.7)$$

Preuve :

$$\tau_l^2 = \sigma^2(F.1_{T > l} - \mathbb{E}F.1_{T > l}) = Var(F.1_{T > l}) + 2 \sum_{k=1}^{\infty} Cov(F.1_{T > l} \circ \theta^k, F.1_{T > l}).$$

Par la même démonstration que le lemme précédent on obtient :

$$|Cov(F.1_{T > l} \circ \theta^k, F.1_{T > l})| \leq \int_{[0,1]^N} |F|^2 d\lambda \frac{(\mathbb{E}T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}}$$

D'où τ_l^2 est une série absolument convergente uniformément en l . Puisque chaque terme converge vers 0, $\lim_{l \rightarrow \infty} \tau_l^2 = 0$. ■

Démontrons maintenant le théorème.

Preuve : On se donne $F \in L^2([0,1]^N, \mathcal{F}_T)$, pour tout $l \in \mathbb{N}$ on écrit

$$F = (F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}) + (F.1_{T > l} - \mathbb{E}F.1_{T > l}).$$

Par suite

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}) \circ \theta^n + \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T > l} - \mathbb{E}F.1_{T > l}) \circ \theta^n.$$

La fonction $F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}$ est $\sigma(U_1, \dots, U_l)$ -mesurable. D'après la proposition 4, le premier terme à droite de l'égalité converge en loi vers $\mathcal{N}(0, \sigma_l^2)$, pour tout $l \in \mathbb{N}$, avec

$$\sigma_l^2 = \sigma^2(F.1_{T \leq l}) = Var(F.1_{T \leq l}) + 2 \sum_{k=1}^{\infty} Cov(F.1_{T \leq l} \circ \theta^k, F.1_{T \leq l}).$$

D'après le lemme 3, on a la convergence en loi de $\mathcal{N}(0, \sigma_l^2)$ vers $\mathcal{N}(0, \sigma^2)$. Par conséquent, il suffit de montrer que :

$$\lim_{l \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T>l} - \mathbb{E}F.1_{T>l}) o\theta^n \right| \geq \varepsilon \right\} = 0$$

pour tout ε positif. D'après l'inégalité de Bienaymé-Tchebichev

$$\mathbb{P} \left\{ \left| \frac{1}{\sqrt{N}} \sum_{n=0}^N (F.1_{T>l} - \mathbb{E}F.1_{T>l}) o\theta^n \right| \geq \varepsilon \right\} \leq \frac{\int_{[0,1]^{\mathbb{N}}} \left| \sum_{n=0}^N (F.1_{T>l} - \mathbb{E}F.1_{T>l}) o\theta^n \right|^2}{N\varepsilon^2}.$$

En utilisant la proposition 3 (dans le chapitre 3) et en faisant tendre N vers l'infini, il résulte que :

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T>l} - \mathbb{E}F.1_{T>l}) o\theta^n \right| \geq \varepsilon \right\} \leq \frac{\tau_l^2}{\varepsilon^2}.$$

La démonstration est achevée par simple application du lemme 4. ■

De la même façon, on démontre le théorème de la limite centrale pour les fonctionnelles F , \mathcal{F}_T -mesurable de moment q , $q > 2$, avec T un temps d'arrêt vérifiant $\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{p}} < \infty$ où $\frac{1}{p} + \frac{1}{q} = 1$.

Théorème 9 Pour toute $F \in L^q([0, 1]^{\mathbb{N}}, \mathcal{F}_T)$, $q \geq 2$, telle que $\int_{[0,1]^{\mathbb{N}}} F d\lambda = 0$ et telle que T vérifie $\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{p}} < \infty$ avec $\frac{1}{p} + \frac{1}{q} = 1$, on a

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1),$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite.

Preuve : On utilise la même démonstration que pour le théorème précédent. On décompose

la somme $\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n$. Pour tout $l \in \mathbb{N}$ on a :

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}) o\theta^n + \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T > l} - \mathbb{E}F.1_{T > l}) o\theta^n.$$

Le premier terme à droite de l'égalité converge en loi vers $\mathcal{N}(0, \sigma_l^2)$, pour tout $l \in \mathbb{N}$, avec

$$\sigma_l^2 = \sigma^2(F.1_{T \leq l}) = \text{Var}(F.1_{T \leq l}) + 2 \sum_{k=1}^{\infty} \text{Cov}(F.1_{T \leq l} o\theta^k, F.1_{T \leq l}).$$

Le deuxième terme, lui, vérifie :

$$\limsup_{N \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (F.1_{T > l} - \mathbb{E}F.1_{T > l}) o\theta^n \right| \geq \varepsilon \right\} \leq \frac{\tau_l^2}{\varepsilon^2}.$$

Il suffit, donc, de démontrer que $\lim_{l \rightarrow \infty} \sigma_l^2 = \sigma^2$ et $\lim_{l \rightarrow \infty} \tau_l^2 = 0$. Pour ce faire, on utilise l'inégalité 3.25 démontrée dans le théorème 6 du chapitre précédent, à savoir que :

$$| \langle F \circ \theta^k, F \rangle | \leq \left(\int_{[0,1]^N} |F|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}}.$$

En remplaçant la fonction $F.1_{T \leq l}$, dans cette inégalité, en lieu et place de la fonction F , on majore $|Cov(F.1_{T \leq l} \circ \theta^k, F.1_{T \leq l})|$ par

$$\left(\int_{[0,1]^N} |F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F.1_{T \leq l} - \mathbb{E}F.1_{T \leq l}|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}}.$$

Or pour tous $p > 1$ on a $\|F - \mathbb{E}F\|_p \leq 2\|F\|_p$. Par suite

$$\begin{aligned} |Cov(F.1_{T \leq l} \circ \theta^k, F.1_{T \leq l})| &\leq 4 \left(\int_{[0,1]^N} |F.1_{T \leq l}|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F.1_{T \leq l}|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}} \\ &\leq 4 \left(\int_{[0,1]^N} |F|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}}. \end{aligned}$$

Ainsi, on démontre que la série σ_l^2 est absolument convergente uniformément en l . Puisque chaque terme converge vers $\langle F \circ \theta^k, F \rangle$, $\lim_{l \rightarrow \infty} \sigma_l^2 = \sigma^2$. Par le même procédé, on montre que :

$$|Cov(F.1_{T > l} \circ \theta^k, F.1_{T > l})| \leq 4 \left(\int_{[0,1]^N} |F|^p d\lambda \right)^{\frac{1}{p}} \cdot \left(\int_{[0,1]^N} |F|^q d\lambda \right)^{\frac{1}{q}} \cdot (\mathbb{P}(T > k))^{\frac{1}{p}}$$

et on conclut que $\lim_{l \rightarrow \infty} \tau_l^2 = 0$. ■

En particulier, on peut dire que si le temps d'arrêt T est intégrable et la fonctionnelle F est bornée alors le théorème de la limite centrale est vérifié.

Etudions maintenant le cas général lorsque F n'est pas nécessairement \mathcal{F}_T -mesurable. Cependant, on suppose que l'on arrive à approcher la fonctionnelle F , à une vitesse suffisante par une suite de fonctions $(F_k)_{k \in \mathbb{N}}$ ne dépendant que d'un nombre fini de coordonnées. Nous les appellerons les fonctions pratiquement cylindriques.

4.2 Théorème de la limite centrale pour les fonctions pratiquement cylindriques.

Dans cette section, on démontre, par les mêmes techniques que le paragraphe précédent, le théorème de la limite centrale pour la classe de fonctions F , introduites au troisième chapitre, et vérifiant : on suppose qu'il existe une suite de fonctions $(F_k)_{k \in \mathbb{N}}$ dans $L^2([0,1]^N)$, telle que F_k est $\sigma(U_1, \dots, U_k)$ -mesurable et que F_k converge vers F dans $L^2([0,1]^N, \mathcal{B}([0,1]^{\otimes N}), \lambda, \theta)$ avec $\sum_{k=1}^{\infty} \|F - F_k\|_2 < \infty$. Lorsque F est \mathcal{F}_T -mesurable on prenait $F_k = F.1_{T \leq k}$. Dans le cas général les fonctions F_k ne sont pas explicite. Cependant on peut les prendre égales à $\mathbb{E}(F/\mathcal{F}_k)$. En effet, F_k étant \mathcal{F}_k -mesurable, il résulte de la définition même de l'espérance conditionnelle que :

$$\|F - \mathbb{E}(F/\mathcal{F}_k)\|_2 \leq \|F - F_k\|_2.$$

Il n'y a donc aucune perte de généralité de supposer $F_k = \mathbb{E}(F/\mathcal{F}_k)$.

Théorème 10 Soit F une fonctionnelle d'intégrale nulle et ayant un moment d'ordre 2. Si l'on suppose qu'il existe $(F_k)_{k \in \mathbb{N}}$ une suite de $L^2([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$, telle que F_k est $\sigma(U_1, \dots, U_k)$ -mesurable et que F_k converge vers F dans $L^2([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$ avec $\sum_{k=1}^{\infty} \|F - F_k\|_2 < \infty$ et si $\sigma^2 := \sigma^2(F) > 0$ alors

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1), \quad (4.8)$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite.

Preuve : On peut supposer que $F_k = \mathbb{E}(F/\mathcal{F}_k)$. Notons que $(F_k)_{k \in \mathbb{N}}$ est alors une $(\mathcal{F}_k)_{k \in \mathbb{N}}$ -martingale. On pose $R_k = F - F_k$ alors $\mathbb{E}F_k = \mathbb{E}R_k = 0$. Pour tous $l \in \mathbb{N}$, on a la décomposition suivante :

$$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F_l \circ \theta^n + \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} R_l \circ \theta^n.$$

L'idée est la même : Approcher le premier terme à droite de l'égalité par une normale lorsque N est grand et montrer que le deuxième terme est petit. En effet, d'après la proposition 4,

$\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} F_l \circ \theta^n$ converge en loi vers une normale $\mathcal{N}(0; \sigma_l^2)$, avec

$$\sigma_l^2 = \sigma^2(F_l) = \text{Var}(F_l) + 2 \sum_{k=1}^{\infty} \text{Cov}(F_l \circ \theta^k, F_l).$$

Maintenant, il faut démontrer que $\lim_{l \rightarrow \infty} \sigma_l^2 = \sigma^2$. Or, d'après l'inégalité 3.19 on a, pour tout $l \in \mathbb{N}$:

$$| \langle F_l \circ \theta^k, F_l \rangle | \leq \|F_l\|_2 \|F_l - \mathbb{E}(F_l/\mathcal{F}_k)\|_2 = \|F_l\|_2 \|F_l - F_{l \wedge k}\|_2.$$

On veut majorer $\|F_l\|_2 \|F_l - F_{l \wedge k}\|_2$ indépendamment de l . L'inégalité de Jensen nous donne $\|F_l\|_2 \leq \|F\|_2$. Outre et, si $k < l$,

$$\|F_l - F_k\|_2 = \|\mathbb{E}(F - F_k/\mathcal{F}_l)\|_2 \leq \|F - F_k\|_2.$$

Si $k \leq l$, le second membre est nul. Donc σ_l^2 est une série absolument convergente uniformément en l . Puisque chaque terme de la série converge vers $\text{Cov}(F \circ \theta^k, F)$, $\lim_{l \rightarrow \infty} \sigma_l^2 = \sigma^2$.

D'après la démonstration du théorème 8, il suffit de vérifier que $\lim_{l \rightarrow \infty} \tau_l^2 = 0$, avec

$$\tau_l^2 := \sigma^2(R_l) = \text{Var}(R_l) + 2 \sum_{k=1}^{\infty} \text{Cov}(R_l \circ \theta^k, R_l).$$

D'après l'inégalité 3.19 on a maintenant :

$$| \langle R_l \circ \theta^k, R_l \rangle | \leq \|R_l\|_2 \|R_l - \mathbb{E}(R_l/\mathcal{F}_k)\|_2.$$

D'une part $\|R_l\|_2 = \|F - F_l\|_2 \leq \|F\|_2 + \|F_l\|_2 \leq 2\|F\|_2$. D'autre part :

$$\begin{aligned} \|R_l - \mathbb{E}(R_l/\mathcal{F}_k)\|_2 &= \|F - F_l - \mathbb{E}(F/\mathcal{F}_k) + \mathbb{E}(F_l/\mathcal{F}_k)\|_2 \\ &\leq \|F - F_k\|_2 + \|F_l - \mathbb{E}(F_l/\mathcal{F}_k)\|_2 \\ &\leq 2\|F - F_k\|_2. \end{aligned}$$

Par suite τ_l^2 est une série absolument convergente uniformément en l . Chaque terme converge vers 0, donc, $\lim_{l \rightarrow \infty} \tau_l^2 = 0$. ■

Il résulte du deuxième exemple de la page 33 et du théorème précédent que les fonctions lipschitziennes vérifient le théorème de la limite centrale.

Corollaire 4 *Soit F une fonction lipschitzienne sur $[0, 1]^N$ d'intégrale nulle. Si l'on suppose que $\sigma^2 := \sigma^2(F) > 0$ alors*

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1), \quad (4.9)$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite. ■

Dans le paragraphe suivant on démontre ce dernier résultat par des techniques différentes, issues de l'Analyse fonctionnelle. Ces techniques fournissent également un résultat de type Berry-Essen.

4.3 Théorème de la limite centrale pour les fonctions lipschitziennes.

Dans ce paragraphe, on utilisera d'autres techniques d'analyse fonctionnelle pour obtenir le théorème de la limite centrale pour la classe des fonctions lipschitziennes. En étudiant le spectre de l'opérateur de Perron-Frobenius associé au shift θ , on montre que σ^2 étant fini et non nul, alors le théorème de la limite centrale est vérifié. Cette technique a été utilisée par Doeblin et Fortet qui ont en particulier étudié les deux exemples principaux parmi les transformations dilatantes et monotones par morceaux sur l'intervalle unité $[0, 1]$: la transformation "fraction continue" et la transformation " $(2x) \bmod 1$ ", voir Rousseau ([48]). Comme il a été fait dans cet article ([48]) et celui de Guivarch et Hardy ([21]), par ces techniques on peut aller plus loin que le théorème de la limite centrale et obtenir une vitesse en $\frac{1}{\sqrt{n}}$ de ce théorème. On démontrera ce résultat dans le cas du shift.

4.3.1 L'opérateur de Perron-Frobenius et ses perturbations.

Reprenons le système dynamique $([0, 1]^N, \mathcal{B}([0, 1]^N), \lambda, \theta)$. L'opérateur de Perron-Frobenius associé à la fonction de décalage θ est l'opérateur P de $L^1([0, 1]^N, \lambda)$ dans $L^1([0, 1]^N, \lambda)$ défini

par :

$$\int_{[0,1]^N} P(F) \cdot G d\lambda = \int_{[0,1]^N} F \cdot G \circ \theta d\lambda$$

où $F \in L^1([0,1]^N)$ et $G \in L^\infty([0,1]^N)$. La définition de l'opérateur θ permet de donner une forme explicite à P :

$$PF(U) = \int_0^1 F(x, U) dx.$$

On notera $C([0,1]^N)$ et $Lip([0,1]^N)$ respectivement les espaces des fonctions continues et des fonctions lipschitziennes et on les normera respectivement par $\|F\|_\infty = \sup_{U \in [0,1]^N} |F(U)|$

et $\|F\|_{Lip} = [F] + \|F\|_\infty$ avec $[F] = \sup_{U, U' \in [0,1]^N} \frac{|F(U) - F(U')|}{d(U, U')}$ et $d(U, U') = \sum_{k=1}^{\infty} \frac{|U_k - U'_k|}{2^k}$, cette dernière distance définie la topologie produit sur $[0,1]^N$.

Rappelons dans un premier temps certains éléments de la théorie spectrale. Soient E un espace de Banach sur \mathbb{C} et ϕ un opérateur linéaire de E dans E alors le spectre de E est défini comme le complémentaire de l'ensemble résolvant.

Définition 4 On appelle ensemble résolvant de ϕ , noté $\mathcal{R}(\phi)$, l'ensemble :

$$\mathcal{R}(\phi) = \{z \in \mathbb{C} \text{ tel que } (zI - \phi)^{-1} \text{ existe}\}.$$

L'ensemble complémentaire est appelé spectre de ϕ et il est noté $\sigma(\phi)$.

On montre que $\mathcal{R}(\phi)$ est un ouvert et que $\sigma(\phi)$ est un compact non vide (voir par exemple [14]). Le rayon spectrale d'un opérateur linéaire est défini par :

Définition 5 On appelle rayon spectrale de ϕ de réel positif $\rho(\phi)$ définie par :

$$\rho(\phi) = \sup_{z \in \sigma(\phi)} |z|.$$

On montre que $\rho(\phi) = \lim_{n \rightarrow \infty} \|\phi^n\|^{\frac{1}{n}}$ (voir par exemple [14]).

Etudions maintenant le spectre de P , P étant ici considéré comme un opérateur sur le sous-espace des fonctions lipschitziennes. Il est clair que $Lip([0,1]^N, \|\cdot\|_{Lip})$ est un espace de Banach et que $Lip([0,1]^N)$ est dense dans $C([0,1]^N, \|\cdot\|_\infty)$. Le spectre de P est décrit à l'aide du théorème de Ionescu-Tulcea et Marinescu ([29],[37]) :

Théorème 11 Soient \mathcal{V} et \mathcal{L} deux espaces de Banach complexes de normes respectives $\|\cdot\|_{\mathcal{V}}$ et $\|\cdot\|_{\mathcal{L}}$ avec $\mathcal{V} \subset \mathcal{L}$. On suppose :

- (a) Si $f_n \in \mathcal{V}$, $f \in \mathcal{L}$, $\lim_{n \rightarrow \infty} \|f_n - f\|_{\mathcal{L}} = 0$ et $\|f_n\|_{\mathcal{V}} \leq M$ pour tout n , alors $f \in \mathcal{V}$ et $\|f\|_{\mathcal{V}} \leq M$.

Soit ϕ un opérateur borné de \mathcal{V} dans \mathcal{L} , par rapport à la norme $\|\cdot\|_{\mathcal{V}}$. On suppose de plus :

- (b) $\sup_{n \geq 0} \{\|\phi^n f\|_{\mathcal{L}}, f \in \mathcal{V}, \|f\|_{\mathcal{L}} \leq 1\} < \infty$.

(c) L'inégalité de Doeblin et Fortet est vérifiée, c'est-à-dire, il existe n_0 , $\alpha < 1$ et $\beta < \infty$ tel que :

$$\|\phi^{n_0} f\|_{\mathcal{V}} < \alpha \|f\|_{\mathcal{V}} + \beta \|f\|_{\mathcal{L}} \quad \text{pour tout } f \in \mathcal{V}.$$

(d) Si V est une partie bornée de $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ alors $\phi^{n_0} V$ est relativement compacte dans $(\mathcal{L}, \|\cdot\|_{\mathcal{L}})$.

Alors ϕ n'a qu'un nombre fini de valeurs propres de modules 1 : $\lambda_1, \dots, \lambda_p$. Les sous-espaces propres correspondants $E_i = \{f \in \mathcal{L} : \phi f = \lambda_i f\}$ sont de dimension finie et contenus dans \mathcal{V} . Ainsi l'espace \mathcal{L} se décompose en somme directe des sous-espaces E_i plus un espace supplémentaire à $\bigoplus_{i=1}^p E_i$:

$$\mathcal{L} = \bigoplus_{i=1}^p E_i \oplus F. \quad (4.10)$$

Soient ϕ_i les projections sur les sous-espaces propres E_i , où les projections se faisant parallèlement au supplémentaire dans la décomposition (4.10) ci-dessus. Ces opérateurs vérifient $\|\phi_i\|_{\mathcal{L}} \leq 1$, $\phi_i \phi_j = \phi_j \phi_i = 0$ si $i \neq j$ et $\phi_i^2 = \phi_i$. L'opérateur ϕ peut s'écrire :

$$\phi = \sum_{i=1}^p \lambda_i \phi_i + \psi,$$

où ψ est un opérateur sur \mathcal{L} tel que $\sup_{n \geq 1} \|\psi^n\|_{\mathcal{L}} < \infty$. Il vérifie de plus $\phi_i \psi = \psi \phi_i = 0$, $\forall i \in \{1, \dots, p\}$. Si l'on note par ϕ^n les puissances de ϕ , alors ϕ^n peut s'écrire :

$$\phi^n = \sum_{i=1}^p \lambda_i^n \phi_i^n + \psi^n \quad \forall n \geq 1,$$

Enfin $\psi(\mathcal{V}) \subset \mathcal{V}$ et ψ a un rayon spectrale $\rho(\psi) < 1$ dans $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$. ■

L'opérateur de Perron-Frobenius associé au shift vérifie les hypothèses du théorème.

Proposition 5 L'opérateur P vérifie les hypothèses du théorème précédent avec $\mathcal{V} = Lip([0, 1]^N)$, $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_{\infty} + [\cdot]$, $\mathcal{L} = C([0, 1]^N)$ et $\|\cdot\|_{\mathcal{L}} = \|\cdot\|_{\infty}$.

Preuve : En effet soit F_n une suite de fonctions lipschitziennes, vérifiant $\|F_n\|_{Lip} \leq M$, et F une fonction continue telle que $\lim_{n \rightarrow \infty} \|F_n - F\|_{\infty} = 0$, alors $[F_n] \leq M$ et pour tout U et $U' \in [0, 1]^N$ on a :

$$\forall n \in \mathbb{N} \quad |F_n(U) - F_n(U')| \leq [F_n] d(U, U') \leq M d(U, U').$$

par passage à la limite il résulte que $|F(U) - F(U')| \leq M d(U, U')$ et donc F est lipschitzienne. Il est clair que P est un opérateur de norme plus petite que 1 et par suite les puissances P^n de P sont bornées par 1. La quatrième propriété est un corollaire du théorème d'Ascoli à savoir qu'une famille de fonctions uniformément lipschitzienne étant équicontinue elle est

relativement compacte pour $\|\cdot\|_\infty$ dès qu'elle est bornée. Il nous reste à établir l'inégalité de Doeblin et Fortet. Soit F lipschitzienne alors ;

$$\begin{aligned} |PF(U) - PF(U')| &\leq \int_{[0,1]^N} |F(x, U) - F(x, U')| d\lambda \\ &\leq [F] \sum_{k=1}^{\infty} \frac{|U_k - U'_k|}{2^{k+1}} = \frac{1}{2} [F] d(U, U'). \end{aligned}$$

Il en ressort que PF est lipschitzienne, $[PF] \leq [F]$ et que :

$$\|PF\|_{Lip} \leq \frac{1}{2} \|F\|_{Lip} + \frac{1}{2} \|F\|_\infty.$$

Ceci achève la démonstration. ■

Du fait que θ est fortement mélangeante on démontre dans le lemme suivant que P admet une seule valeur propre de norme 1. Ainsi la valeur propre 1 est seule et isolée sur le cercle de rayon 1.

Lemme 5 Si $PF = e^{i\omega} F$ avec $\omega \in \mathbb{R}$ et $F \in L^1([0,1]^N)$ alors $e^{i\omega} = 1$ et la fonction F est constante.

Preuve : En effet, pour tout $n \in \mathbb{N}$, on a $P^n F = e^{in\omega} F$. Il se déduit que :

$$\forall G \in L^\infty([0,1]^N) \text{ et } n \in \mathbb{N} \quad \int_{[0,1]^N} F \cdot G \circ \theta^n d\lambda = \int_{[0,1]^N} P^n F \cdot G d\lambda = e^{in\omega} \int_{[0,1]^N} F \cdot G d\lambda$$

Lorsque n tend vers l'infini le membre à gauche converge vers $\int_{[0,1]^N} F d\lambda \int_{[0,1]^N} G d\lambda$, puisque θ est fortement mélangeante. Le membre à droite lui converge que pour $\omega = 0$. Par conséquent, on a $\omega = 0$ et F est une fonction constante. ■

Soit F une fonction de $Lip([0,1]^N)$, à valeurs réelles et $\omega \in \mathbb{R}$. On pose $P(F, i\omega)(G) = P(e^{i\omega F} G)$. Cet opérateur s'introduira naturellement dans l'étude de la fonction caractéristique de $\sum_{k=1}^{n-1} F \circ \theta^k$. Une décomposition spectrale de $P(F, i\omega)$ nous permettra de calculer ses itérées.

Dans la proposition suivante nous allons décrire le spectre de $P(F, i\omega)$ lorsque ω est voisin de 0. Ce résultat est dû à Rellich ([14]) qui a décrit comment les points isolés du spectre d'un opérateur varient lorsque l'on fait dépendre cet opérateur d'un paramètre d'une manière analytique. Dans le lemme suivant, on montre que l'opérateur $P(F, i\omega)$ dépend de ω analytiquement.

Lemme 6 Pour tout $\omega \in \mathbb{R}$, l'opérateur $P(F, i\omega)$ est un opérateur continu sur $Lip([0,1]^N)$ et l'application qui à ω fait correspondre $P(F, i\omega)$ est analytique.

Preuve : $P(F, i\omega)$ est le composé d'opérateurs continus. Pour la deuxième affirmation, à savoir que l'application $\omega \longrightarrow P(F, i\omega)$ est analytique, il suffit de remarquer que

$$\frac{|\omega|^k}{k!} \|P(F^k G)\|_{Lip} \leq \frac{|\omega|^k}{k!} \|F\|_{Lip}^k \|G\|_{Lip}.$$

D'où, la série $\sum_{k=0}^{\infty} \frac{(i\omega)^k}{k!} P(F^k G)$ est normalement convergente dans $Lip([0, 1]^N)$.

■

Le résultat suivant nous permet de démontrer le théorème de la limite centrale. On définit par perturbation, au voisinage de $\omega = 0$, une valeur propre dominante $k(F, i\omega)$, également simple et isolée.

Proposition 6 *Il existe un réel $a > 0$ tel que si $|\omega| < a$, on ait :*

(a) *Pour tout $G \in L([0, 1]^N)$ et $n \geq 1$*

$$P^n(F, i\omega)(G) = k(F, i\omega)^n \phi_{i\omega}(G) + \psi_{i\omega}(G)$$

où, $k(F, i\omega)$ est l'unique valeur propre dominante de $P(F, i\omega)$, $\phi_{i\omega}$ est la projection sur le sous espace propre $E_{i\omega}$ correspondant à $k(F, i\omega)$, $\psi_{i\omega}$ est un opérateur sur $Lip([0, 1]^N)$ de rayon spectral

$$\rho(\psi_{i\omega}) \leq 1 - \varepsilon \quad \varepsilon > 0$$

et $\psi_{i\omega} E_{i\omega} = 0$.

(b) *Les applications $\omega \longrightarrow k(F, i\omega)$, $\omega \longrightarrow \phi_{i\omega}$ et $\omega \longrightarrow \psi_{i\omega}$ sont analytiques.*

(c) *$\|\psi_{i\omega}^n(1)\|_{Lip} \leq C|\omega|(1 - \varepsilon)^n$ où C est une constante positive.*

Preuve : Pour la démonstration voir ([14],[48]).

■

La fonction $k(F, i\omega)$ possède des propriétés analogues à celles d'une fonction caractéristique. Dans le lemme suivant on donne un développement limité à l'ordre 2 de la fonction $k(F, i\omega)$ pour ω petit.

Lemme 7 *Soit F une fonction lipschitzienne sur $[0, 1]^N$, $k(F, i\omega)$ la valeur propre dominante de $P(F, i\omega)$ définie ci-dessus. Alors la fonction $k(F, i\omega)$ est analytique au voisinage de 0. En particulier $k(F, 0) = 1$, $k'(F, 0) = \int_{[0, 1]^N} F d\lambda$ et $k''(F, 0) = \sigma^2$ pour les fonctions d'intégrale nulle.*

Preuve : On note par M et Q les opérateurs définis sur $Lip([0, 1]^N)$ par $M(G) = P(F G)$ et $Q(G) = P(F^2 G)$. La série $\sum_{k=0}^{\infty} \frac{(i\omega)^k}{k!} P(F^k G)$ étant normalement convergente on peut dire que :

$$P(F, i\omega) = P + i\omega M - \frac{\omega^2}{2} Q + o(\omega^2). \quad (4.11)$$

D'après la proposition précédente on a :

$$P(F, i\omega) = k(F, i\omega) \phi_{i\omega} + \psi_{i\omega} \quad (4.12)$$

avec $\phi_{i\omega} \psi_{i\omega} = \psi_{i\omega} \phi_{i\omega} = 0$, $\phi_{i\omega} = \phi_{i\omega}^2$ et les fonctions $\omega \longrightarrow \phi_{i\omega}$, $\omega \longrightarrow k(F, i\omega)$ sont analytiques. Un développement limité d'ordre 2 de ces fonctions nous donne : d'une part

$$\phi_{i\omega} = \phi + i\omega \alpha - \frac{\omega^2}{2} \beta + o(\omega^2) \quad (4.13)$$

où ϕ est la projection sur le sous espace propre E_1 correspondant à la valeur propre 1 de l'opérateur $P(F, 0) = P$; d'autre part

$$k(F, i\omega) = 1 + i\omega a - \frac{\omega^2}{2}b + o(\omega^2). \quad (4.14)$$

A partir de ces équations, on peut identifier a et b (définies dans l'équation 4.14) qui représentent la dérivée première et la dérivée seconde de la fonction $k(F, i\omega)$ au point $\omega = 0$. D'une part, le développement limité d'ordre 2 de la fonction $k(F, i\omega)\phi_{i\omega}$:

$$\begin{aligned} k(F, i\omega)\phi_{i\omega} &= (1 + i\omega a - \frac{\omega^2}{2}b + o(\omega^2))(\phi + i\omega\alpha - \frac{\omega^2}{2}\beta + o(\omega^2)) \\ &= \phi + i\omega(a\phi + \alpha) - \frac{\omega^2}{2}(2a\alpha + b\phi + \beta) + o(\omega^2). \end{aligned}$$

D'autre part, on a aussi le développement limité d'ordre 2 de la fonction $\phi_{i\omega}P(F, i\omega)$:

$$\begin{aligned} \phi_{i\omega}P(F, i\omega) &= (\phi + i\omega\alpha - \frac{\omega^2}{2}\beta + o(\omega^2))(P + i\omega M - \frac{\omega^2}{2}Q + o(\omega^2)) \\ &= \phi P + i\omega(\alpha P + \phi M) - \frac{\omega^2}{2}(\phi Q + 2\alpha M + \beta P) + o(\omega^2). \end{aligned}$$

Or, on a vu ci-dessus que $P(F, i\omega) = k(F, i\omega)\phi_{i\omega} + \psi_{i\omega}$, avec $\phi_{i\omega}\psi_{i\omega} = \psi_{i\omega}\phi_{i\omega} = 0$, $\phi_{i\omega} = \phi_{i\omega}^2$ et donc $k(F, i\omega)\phi_{i\omega} = \phi_{i\omega}P(F, i\omega)$. Par identification des développements limités ci-dessus on obtient le système suivant :

$$\begin{cases} \phi P &= \phi \\ a\phi + \alpha &= \alpha P + \phi M \\ \phi Q + 2\alpha M + \beta P &= 2a\alpha + b\phi + \beta \end{cases}$$

En appliquant la deuxième équation du système à la fonction 1, il résulte que

$$a\phi(1) = M\phi(1).$$

Comme $\phi(1) = 1$, on déduit que $a = M\phi(1)$. Par définition de M , $M(1) = P(F)$, on obtient, donc, que a est la projection de PF sur le sous espace propre E_1 associé à la valeur propre 1 de l'opérateur P . D'après le lemme 5, E_1 est de dimension 1 et il est égale à l'ensemble des fonctions constantes. Par conséquent :

$$a = \langle PF, 1 \rangle . 1 = \int_{[0,1]^N} PF d\lambda = \int_{[0,1]^N} F d\lambda.$$

Dans la suite de la démonstration, on suppose que la fonction F est d'intégrale nulle. D'après l'équation ci-dessus on a $a = 0$. La troisième équation du système devient

$$\phi Q + 2\alpha M + \beta P = b\phi + \beta.$$

En l'appliquant à la fonction 1, il résulte que :

$$b = \phi P(F^2) + 2\alpha P(F). \quad (4.15)$$

Comme la fonction F est d'intégrale nulle, la deuxième équation du système s'écrit

$$\alpha(I - P) = \phi M.$$

Or $(I - P)$ est inversible puisque le rayon spectrale de P est strictement plus petit que 1 sur l'ensemble des fonctions lipschitziennes sur $[0, 1]^N$ et d'intégrale nulle. Il ressort que

$$\alpha = \phi M(I + P + P^2 + \dots)$$

et que

$$b = \phi P(F^2) + 2\phi M(I + P + P^2 + \dots)P(F).$$

Le premier terme à droite de l'égalité s'identifie à :

$$\phi P(F^2) = \langle P(F^2), 1 \rangle = \int_{[0,1]^N} P(F^2) d\lambda = \|F\|_2^2,$$

le deuxième terme lui se calcule comme suit :

$$\begin{aligned} 2\phi M(I + P + P^2 + \dots)PF &= 2\phi M(PF + P^2F + \dots) \\ &= 2 \sum_{k=1}^{\infty} \phi P(F P^k F) = 2 \sum_{k=1}^{\infty} \int_{[0,1]^N} F P^k F d\lambda \\ &= 2 \sum_{k=1}^{\infty} \langle F, P^k F \rangle = 2 \sum_{k=1}^{\infty} \langle F \circ \theta^k, F \rangle \end{aligned}$$

Ainsi, $k'(F, 0) = \int_{[0,1]^N} F d\lambda$ et $k''(F, 0) = \sigma^2$ pour les fonctions d'intégrale nulle. ■

4.3.2 Théorème de la limite centrale.

Considérons S_n la somme ergodique associée à la transformation du shift et définie par $S_n = \sum_{k=0}^{n-1} F \circ \theta^k$. Pour montrer le théorème de la limite centrale on calcule la limite de la

fonction caractéristique de $\frac{S_n}{\sqrt{n}}$.

Théorème 12 Soit F une fonction lipschitzienne sur $[0, 1]^N$ d'intégrale nulle et de $\sigma^2(F) > 0$ alors :

$$\frac{1}{\sigma\sqrt{N}} \sum_{n=0}^{N-1} F \circ \theta^n \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1), \quad (4.16)$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite.

Preuve : On note par Φ_n la transformée de Fourier de la loi de $\frac{S_n}{\sqrt{n}}$,

$$\Phi_n(\omega) = \int_{[0,1]^N} e^{i\omega \frac{S_n}{\sqrt{n}}} d\lambda \quad \text{pour tout } n \in \mathbb{N} \text{ et } \omega \in \mathbb{R}.$$

On pose $T(F, i\omega)(G) = e^{i\omega F} (G \circ \theta)$. L'opérateur $P(F, i\omega)$ s'introduit naturellement dans le calcul des Φ_n via $T(F, i\omega)$. En effet, $P(F, i\omega)$ est l'opérateur adjoint de $T(F, i\omega)$ et $T^n(F, i\omega)(G) = e^{i\omega S_n} G \circ \theta^n$. La fonction caractéristique de $\frac{S_n}{\sqrt{n}}$ s'exprime en fonction de l'opérateur $P(F, i\omega)$ de la façon suivante :

$$\begin{aligned}\Phi_n(\omega) &= \langle T^n(F, \frac{i\omega}{\sqrt{n}})1, 1 \rangle \\ &= \langle 1, P^n(F, \frac{i\omega}{\sqrt{n}})1 \rangle.\end{aligned}\quad (4.17)$$

Reprenons la décomposition de $P(F, i\omega)$ donnée dans la proposition 6 (c) : on a pour ω petit $P^n(F, i\omega) = k^n(F, i\omega)\phi_{i\omega} + \psi_{i\omega}^n$. En particulier, on peut écrire que pour $\frac{\omega}{\sqrt{n}}$ petit on a :

$$P^n(F, \frac{i\omega}{\sqrt{n}})1 = k^n(F, \frac{i\omega}{\sqrt{n}})\phi_{\frac{i\omega}{\sqrt{n}}}1 + \psi_{\frac{i\omega}{\sqrt{n}}}^n 1. \quad (4.18)$$

Etudions la convergence des termes à droite de l'égalité 4.18 lorsque n est grand. Il découle du lemme précédent que le développement limité à l'ordre 2 des valeurs propres $k(F, i\omega)$ est donné par :

$$k(F, i\omega) = 1 - \frac{\omega^2}{2}\sigma^2 + o(\omega^2).$$

Ceci montre que $k^n(F, \frac{i\omega}{\sqrt{n}})$ converge vers $e^{-\frac{\omega^2}{2}\sigma^2}$. En remarquant que $\lim_{n \rightarrow \infty} \phi_{\frac{i\omega}{\sqrt{n}}} = \lim_{\omega \rightarrow 0} \phi_{i\omega} = \phi$, on déduit que le premier terme à droite de l'égalité 4.18 converge vers $e^{-\frac{\omega^2}{2}\sigma^2}$. La limite du deuxième terme est donnée par la troisième propriété de la proposition 6 : il existe un réel $a > 0$ tel que si $|\omega| < a$, on ait :

$$\|\psi_{i\omega}^n(1)\|_{Lip} \leq C|\omega|(1-\varepsilon)^n$$

où C est une constante positive. Il résulte que si $\frac{\omega}{\sqrt{n}}$ est petit alors $\|\psi_{\frac{i\omega}{\sqrt{n}}}^n(1)\|_{\infty} \leq C\frac{|\omega|}{\sqrt{n}}(1-\varepsilon)^n$ et donc $\lim_{n \rightarrow \infty} \psi_{\frac{i\omega}{\sqrt{n}}}^n(1) = 0$. On achève la démonstration, par une application du théorème de convergence dominée dans (4.17). Pour ce faire, on montre qu'il existe deux constantes positives A et B telles que pour $\frac{\omega}{\sqrt{n}}$ assez petit la suite $\|P^n(F, \frac{i\omega}{\sqrt{n}})1\|_{\infty}$ soit bornée par $Ae^{-B\omega^2}$. En effet :

$$\|P^n(F, \frac{i\omega}{\sqrt{n}})1\|_{\infty} \leq |k^n(F, \frac{i\omega}{\sqrt{n}})| + \|\psi_{\frac{i\omega}{\sqrt{n}}}^n 1\|_{\infty}.$$

Il est clair que pour $|\omega| \leq \varepsilon'$ on a $|k(F, i\omega)| \leq e^{-\frac{\omega^2}{4}\sigma^2}$ et on déduit que si $\frac{|\omega|}{\sqrt{n}} \leq \varepsilon'$ alors :

$$|k^n(F, \frac{i\omega}{\sqrt{n}})| \leq e^{-\frac{\omega^2}{4}\sigma^2}$$

D'après la propriété 4.3.2 ci-dessus, on déduit que si $\frac{|\omega|}{\sqrt{n}} \leq a$:

$$\|\psi_{\frac{i\omega}{\sqrt{n}}}^n 1\|_{\infty} \leq C\frac{|\omega|}{\sqrt{n}}(1-\varepsilon)^n \leq C a e^{-n\varepsilon} \leq C a e^{-\varepsilon\frac{\omega^2}{a^2}}.$$

Le résultat est obtenu avec $A = C + 1$ et $B = \sup(\frac{\sigma^2}{4}, \frac{\varepsilon}{a^2})$. En vertu de ces résultats, on a la convergence des transformées de Fourier de $\frac{S_n}{\sigma\sqrt{n}}$ vers $e^{-\frac{u^2}{2}}$, ce qui établit le théorème. ■

On veut, maintenant, caractériser les fonctions lipschitziennes vérifiant $\sigma^2 = 0$, en montrant que ce sont des cobords de la transformation du shift.

Proposition 7 *Soit F une fonction lipschitzienne alors $\sigma^2(F) = 0$ si et seulement si F est de la forme $F = G \circ \theta - G$, où G est une fonction lipschitzienne.*

Preuve : Supposons qu'il existe $G \in Lip([0, 1]^N)$ telle la fonction F s'écrit $F = G \circ \theta - G$ alors $\sum_{k=0}^{n-1} F \circ \theta^k = G \circ \theta^n - G$ et par suite :

$$\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{n} Var \left(\sum_{k=0}^{n-1} F \circ \theta^k \right) \leq 2 \lim_{n \rightarrow \infty} \frac{1}{n} Var(G) = 0,$$

Inversement, supposons que $\sigma^2 = 0$. D'après la proposition 6, $P = \phi + \psi$ avec ϕ la projection sur l'espace propre E_1 et ψ est un opérateur de rayon spectrale plus petit que 1, donc $(I - P)$ est inversible sur l'ensemble des fonctions lipschitziennes. En posant $G = \sum_{k=1}^{\infty} P^k F$, G est solution de l'équation de Poisson :

$$(I - P)G = PF, \quad (4.19)$$

et par suite $F = G \circ \theta - G$, puisque

$$\begin{aligned} \|F + G - G \circ \theta\|_2^2 &= \|F + G\|_2^2 + \|G\|_2^2 - 2 \langle G \circ \theta, F + G \rangle \\ &= \|F\|_2^2 + 2 \langle F, G \rangle + 2\|G\|_2^2 - 2 \langle G, PF + PG \rangle \\ &= \|F\|_2^2 + 2 \langle F, G \rangle + 2 \langle G, G - PF - PG \rangle \end{aligned}$$

D'après l'équation 4.19 le troisième terme à droite est nulle. En remarquant que $\sigma^2(F)$ est définie par $\sigma^2(F) := \|F\|_2^2 + 2 \langle F, G \rangle$ on obtient $\|F + G - G \circ \theta\|_2^2 = 0$. Ceci achève la démonstration. ■

Remarque :

Soit G une fonction mesurable, solution de l'équation fonctionnelle :

$$F = G \circ \theta - G.$$

Alors on a $\sum_{k=0}^{n-1} F \circ \theta^k = G \circ \theta^n - G$. Soit $c > 0$, alors :

$$\lambda \left(\frac{G \circ \theta^n}{\sqrt{n}} > c \right) = \lambda \left(\frac{G}{\sqrt{n}} > c \right) \quad (4.20)$$

et donc $\frac{\sum_{k=0}^{n-1} F \circ \theta^k}{\sqrt{n}} = \frac{G \circ \theta^n - G}{\sqrt{n}}$ tend vers 0 en probabilité.

Dans le paragraphe suivant on obtient un résultat plus précis, en démontrant le théorème de la limite centrale avec reste.

4.3.3 La vitesse dans le théorème de la limite centrale.

La méthode précédente permet de préciser la vitesse de convergence dans le théorème de la limite centrale. La démonstration repose sur l'inégalité d'Essen ([16],[50]) et un calcul de développement limité plus poussé que précédemment. Ainsi, on obtient une vitesse exacte en $\frac{1}{\sqrt{n}}$.

Théorème 13 Soit F une fonction lipschitzienne sur $[0, 1]^N$ d'intégrale nulle et de $\sigma^2(F) > 0$ alors il existe une constante $C > 0$ telle que pour tout $t \in \mathbb{R}$, on ait :

$$|\lambda\{\frac{1}{\sigma\sqrt{n}} \sum_{k=0}^{n-1} F \circ \theta^k \leq t\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx| \leq \frac{C}{\sqrt{n}}. \quad (4.21)$$

Preuve : Pour alléger les notations, on prendra $\sigma = 1$. D'après l'inégalité d'Essen, on a pour tout $a > 0$ et $n \geq 1$:

$$\sup_{t \in \mathbb{R}} |\lambda\{\frac{1}{\sigma\sqrt{n}} \sum_{k=0}^{n-1} F \circ \theta^k \leq t\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx| \leq \frac{k}{a} + \frac{1}{\pi} \int_{-a}^a \frac{|\Phi_n(\omega) - e^{-\frac{\omega^2}{2}}|}{|\omega|} d\omega, \quad (4.22)$$

avec $k = \frac{12}{\pi}$ et Φ_n la transformée de Fourier de la loi de $\frac{S_n}{\sqrt{n}}$, $\Phi_n(\omega) = \int_{[0,1]^N} e^{i\omega \frac{S_n}{\sqrt{n}}} d\lambda$. Pour établir le résultat on démontre qu'il existe $A > 0$ telle que pour $\frac{\omega}{\sqrt{n}}$ assez petit, autrement dit il existe $\varepsilon' > 0$ telle que pour $\frac{|\omega|}{\sqrt{n}} < \varepsilon'$, on a :

$$|\Phi_n(\omega) - e^{-\frac{\omega^2}{2}}| \leq A \left(\frac{|\omega|}{\sqrt{n}} (1 - \varepsilon)^n + e^{-\frac{\omega^2}{4}} \left(\frac{|\omega|^3}{\sqrt{n}} + \frac{|\omega|}{\sqrt{n}} \right) \right). \quad (4.23)$$

L'inégalité 4.23 étant supposée valable pour $\frac{\omega}{\sqrt{n}}$ plus petit que ε' , prenant $a = \varepsilon' \sqrt{n}$ dans l'inégalité de Berry-Essen 4.22, le premier terme s'écrit $\frac{12}{\pi \varepsilon' \sqrt{n}}$ et le second terme lui est majoré par :

$$\frac{A}{\pi} \left\{ 2(1 - \varepsilon)^n + \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} (1 + |\omega|^2) e^{-\frac{\omega^2}{4}} d\omega \right\} \quad (4.24)$$

et donc par $\frac{Cte}{\sqrt{n}}$. Ceci donne la majoration voulue. Il reste à établir l'équation 4.23 et donc à majorer $|\Phi_n(\omega) - e^{-\frac{\omega^2}{2}}|$. Comme :

$$|\Phi_n(\omega) - e^{-\frac{\omega^2}{2}}| = \left| \int_{[0,1]^N} e^{i\omega \frac{S_n}{\sqrt{n}}} - e^{-\frac{\omega^2}{2}} d\lambda \right|$$

$$\begin{aligned}
&= \left| \int_{[0,1]^N} P^n(F, i \frac{\omega}{\sqrt{n}}) 1 - e^{-\frac{\omega^2}{2}} d\lambda \right| \\
&\leq \int_{[0,1]^N} |P^n(F, i \frac{\omega}{\sqrt{n}}) 1 - e^{-\frac{\omega^2}{2}}| d\lambda,
\end{aligned}$$

on estime la dernière intégrale à l'aide d'un développement limité, poussé à l'ordre 3 de $P^n(F, i \frac{\omega}{\sqrt{n}})$. On rappelle que :

$$P^n(F, i \frac{\omega}{\sqrt{n}}) = k^n(F, i \frac{\omega}{\sqrt{n}}) \phi_{i \frac{\omega}{\sqrt{n}}} + \psi_{i \frac{\omega}{\sqrt{n}}}^n,$$

avec $\|\psi_{i \frac{\omega}{\sqrt{n}}}^n(1)\|_{Lip} \leq C \frac{|\omega|}{\sqrt{n}} (1 - \varepsilon)^n$ pour $\frac{|\omega|}{\sqrt{n}}$ assez petit. Il résulte que :

$$\begin{aligned}
|P^n(F, i \frac{\omega}{\sqrt{n}}) 1 - e^{-\frac{\omega^2}{2}}| &\leq |k^n(F, i \frac{\omega}{\sqrt{n}}) - e^{-\frac{\omega^2}{2}}| + \\
&|k^n(F, i \frac{\omega}{\sqrt{n}})| \|\phi_{i \frac{\omega}{\sqrt{n}}} 1 - 1\|_{\infty} + \|\psi_{i \frac{\omega}{\sqrt{n}}}^n 1\|_{\infty}.
\end{aligned} \tag{4.25}$$

On constate que, pour $\frac{|\omega|}{\sqrt{n}}$ assez petit, $|k^n(F, i \frac{\omega}{\sqrt{n}})| < e^{-\frac{\omega^2}{4}}$ et que $\|\phi_{i \frac{\omega}{\sqrt{n}}} 1 - 1\|_{Lip} \leq C \frac{|\omega|}{\sqrt{n}}$.

Soit encore il existe $\varepsilon' > 0$ telle que pour $\frac{|\omega|}{\sqrt{n}} < \varepsilon'$ on ait :

$$\begin{aligned}
|P^n(F, i \frac{\omega}{\sqrt{n}}) 1 - e^{-\frac{\omega^2}{2}}| &\leq |k^n(F, i \frac{\omega}{\sqrt{n}}) - e^{-\frac{\omega^2}{2}}| + \\
&C \frac{|\omega|}{\sqrt{n}} e^{-\frac{\omega^2}{4}} + C \frac{|\omega|}{\sqrt{n}} (1 - \varepsilon)^n
\end{aligned} \tag{4.26}$$

Pour majorer $|k^n(F, i \frac{\omega}{\sqrt{n}}) - e^{-\frac{\omega^2}{2}}|$ par $C \frac{|\omega|^3}{\sqrt{n}} e^{-\frac{\omega^2}{4}}$, on utilisera un développement d'ordre 3, de la fonction $k(F, i\omega)$. En effet :

$$k(F, i\omega) = 1 - \frac{\omega^2}{2} - i \frac{\omega^3}{6} \eta + \omega^3 \varepsilon(\omega),$$

avec $\eta = k'''(F, 0)$ et $\lim_{\omega \rightarrow 0} \varepsilon(\omega) = 0$. On a donc :

$$k^n(F, i \frac{\omega}{\sqrt{n}}) = e^{-\frac{\omega^2}{2} - i \frac{\omega^3}{6\sqrt{n}} \eta + \frac{\omega^3}{\sqrt{n}} \varepsilon(\frac{\omega}{\sqrt{n}})}.$$

Soit alors :

$$|k^n(F, i \frac{\omega}{\sqrt{n}}) - e^{-\frac{\omega^2}{2}}| = e^{-\frac{\omega^2}{2}} |e^{i \frac{\omega^3}{6\sqrt{n}} \rho + \frac{\omega^3}{\sqrt{n}} \varepsilon(\frac{\omega}{\sqrt{n}})} - 1|,$$

avec $\rho \in \mathbb{R}$. En outre, on peut trouver un réel $\varepsilon' > 0$ tel que $2|\rho|\varepsilon' < \frac{1}{4}$ et pour $\frac{|\omega|}{\sqrt{n}} < \varepsilon'$ on a :

$$|i \frac{\omega^3}{\sqrt{n}} \rho + \frac{\omega^3}{\sqrt{n}} \varepsilon(\frac{\omega}{\sqrt{n}})| \leq 2\rho \frac{|\omega|^3}{\sqrt{n}} \leq \frac{\omega^2}{4}.$$

On obtient le résultat à l'aide de l'inégalité : $|e^z - 1| \leq |z|e^{|z|}$. ■

Remarque :

La majoration de $|k^n(F, i\frac{\omega}{\sqrt{n}}) - e^{-\frac{\omega^2}{2}}|$ par $A\frac{|\omega|^3}{\sqrt{n}}e^{-\frac{\omega^2}{4}}$ est un résultat élémentaire valable pour toute fonction $h(\omega)$ trois fois dérivable pour ω petit avec $h(0) = 1$, $h'(0) = 0$ et $h''(0) = -1$ [50]. Sous ces hypothèses il existe une constante A telle que :

$$|h^n(\frac{\omega}{\sqrt{n}}) - e^{-\frac{\omega^2}{2}}| \leq A\frac{|\omega|^3}{\sqrt{n}}e^{-\frac{\omega^2}{4}}$$

pour $\frac{\omega}{\sqrt{n}}$ assez petit.

4.4 Illustration informatique.

Comme dans le cas de la convergence presque sûre, nous allons vérifier la validité des résultats théoriques obtenus. Ainsi, on considère des marches aléatoires définies par :

$$X_0 = x, \quad X_{n+1} = X_n + h(U_{n+1}^1, \dots, U_{n+1}^d),$$

où les (U_n^i) sont des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. Pour chaque marche aléatoire, on définit le temps d'arrêt $T = \inf\{n \geq 1; X_n \notin [a, b]\}$ et des fonctionnelles de type $F = G(X_T, T)$. Dans le plupart des applications, T admet un moment d'ordre $p > 2$.

Comme nous l'avons dit au chapitre précédent, l'estimation de σ s'obtient à partir de l'identité (3.18) établie dans la proposition 3, à savoir que :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Var}(\sum_{n=0}^{N-1} F \circ \theta^n) = \sigma^2.$$

Rappelons qu'une stabilisation numérique de cette série sera assimilée à sa valeur limite : ainsi, si l'on note par N_{max} la valeur à partir de laquelle les premiers chiffres des termes de la suite ne sont plus modifiés, on prendra comme estimation de σ la quantité :

$$\hat{\sigma} = \frac{1}{N_{max}} \text{Var}(\sum_{n=0}^{N_{max}-1} F \circ \theta^n).$$

Pour N fixé on simule $\text{Var}(\sum_{n=0}^{N-1} F \circ \theta^n)$ par la méthode de Monte Carlo sur 5 000 trajectoires indépendantes.

Pour observer numériquement la convergence en loi du processus

$$A_n(F) = \frac{F + F \circ \theta + \dots + F \circ \theta^{n-1} - n\mathbb{E}F}{\hat{\sigma}\sqrt{n}},$$

vers la loi normale centrée réduite, on a calculé la distance du χ^2 entre $A_n(F)$ et $\mathcal{N}(0;1)$: ainsi, on a partagé l'espace des observations \mathbb{R} en $m = 57$ catégories par discrétisation de

l'intervalle $] -2.8, 2.8[$, avec des pas de longueur 0.1, et en considérant la catégorie résiduelle $] -\infty, -2.8[\cup] 2.8, +\infty[$. A partir, de $N = 5\,000$ simulations indépendantes de $A_n(F)$ on a calculé les fréquences empiriques $(\frac{n_i}{N})_{1 \leq i \leq m}$. On rappelle que la “distance de Khi-deux” entre $A_n(F)$ et $\mathcal{N}(0; 1)$ (pour les m classes définies en précédemment) est donnée par la formule :

$$D_n^N = D_N(A_n(F), \mathcal{N}(0; 1)) = \sum_{i=1}^m \frac{(n_i - Np_i)^2}{Np_i}, \quad (4.27)$$

où $p_i = \mathbb{P}\{\mathcal{N}(0; 1) \in \text{classe } i\}$.

On sait que lorsque N est grand, on peut approcher la loi de la distance D_n^N par la loi de $\chi^2(m-1)$. Intuitivement, si $A_n(F) \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1)$, D_n^N doit être “petit” pour n grand.

Supposons que l'on veuille savoir si, pour n donné, les simulations de $A_n(F)$ s'assimilent bien à la loi $\mathcal{N}(0; 1)$. Nous utiliserons le test de χ^2 , en tant que test d'adéquation, dont la formulation est la suivante :

$$\begin{cases} (H_0) & \text{la loi de } A_n(F) \text{ est } \mathcal{N}(0; 1) \\ (H_1) & \text{la loi de } A_n(F) \text{ n'est pas } \mathcal{N}(0; 1). \end{cases}$$

La réponse au problème du test est définie par la région critique :

$W = \{(n_1, \dots, n_m) \text{ tel que } D_n^N > C_\alpha\}$, dont la puissance est égale à $\mathbb{P}(W) = \alpha$.

Pour une puissance $\alpha = 0.05$ et $m = 57$, on trouve $C_\alpha = 74.12$. On qualifiera donc la distance D_n^N de “petite” si elle est inférieure à la valeur C_α trouvée.

Exemples : Considérons la marche aléatoire :

$$X_0 = 0, X_{n+1} = X_n + \frac{2U_{n+1} - 1}{2} \text{ et } T = \inf\{n \geq 1, X_n \notin [-5, 1]\}.$$

- Pour $F = X_T$:

En prenant $N_{max} = 10\,000$, on obtient $\hat{\sigma} = 23.724$.

Le tableau suivant représente la variation de D_n^N en fonction de n .

n	100	1 000	5 000	10 000	100 000
D_n^N	31 667.18	759.11	113.92	70.39	67.66

Tableau 4.1

A partir de $n = 10\,000$ la loi de $A_n(X_T)$ s'assimile bien à une gaussienne centrée réduite. La représentation graphique ci-dessous illustre la convergence en loi de $A_n(X_T)$ vers la loi normale $\mathcal{N}(0; 1)$, elle représente l'histogramme associé à $A_{100\,000}(X_T)$.

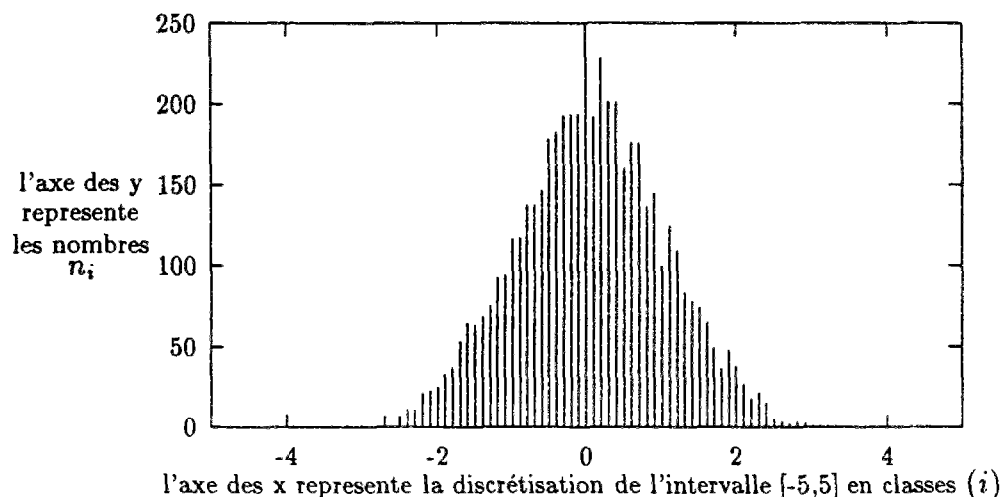


Figure 4.1

- Pour $F = T$: En prenant $N_{max} = 10\,000$, on obtient $\hat{\sigma} = 574.281$. Le tableau suivant représente la variation de D_n^N en fonction de n .

n	100	1 000	5 000	10 000	100 000
D_n^N	2 187.10	258.07	92.10	64.81	72.89

Tableau 4.2

A partir de $n = 10\,000$ la loi de $A_n(T)$ s'assimile bien à une gaussienne centrée réduite. La représentation graphique ci-dessous illustre la convergence en loi de $A_n(T)$ vers la loi normale $\mathcal{N}(0; 1)$, elle représente l'histogramme associé à $A_{100\,000}(T)$.

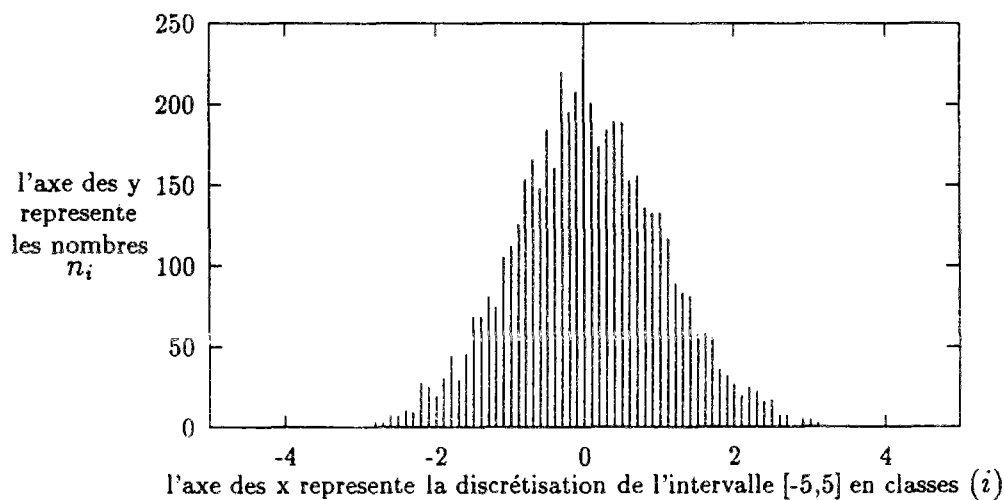


Figure 4.2

4.5 Conclusion.

Les résultats établis ci-dessus montrent que la méthode du shift pour le calcul de l'espérance de fonctionnelles de processus aléatoires vérifie le théorème de la limite centrale pour une large classe de fonctionnelles parmi les plus couramment rencontrées. De plus, la convergence de cette méthode est contrôlée par le paramètre σ^2 . Ce paramètre joue un rôle similaire à celui que joue la variance dans le cas de la méthode de Monte Carlo.

Chapitre 5

Sur la loi du logarithme itéré

Dans le chapitre 3, on montre que pour certaines classes de fonctions l'erreur théorique d'intégration par la méthode du décalage est en $o(N^{-\frac{1}{2}}(\log(N))^{\frac{3}{2}+\varepsilon})$. Ce résultat donne une estimation assez voisine du logarithme itéré; elle est donc suffisante pour la simulation puisqu'elle affirme que l'erreur asymptotique "tourne autour" de $0(\frac{1}{\sqrt{N}})$. Cependant, pour avoir une vitesse de convergence plus fine, j'ai cherché à établir dans ce cadre la loi du logarithme itéré. Le premier paragraphe est consacré à l'étude de la transformation de décalage naturel, donc au système dynamique $([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta)$. Dans ce paragraphe, j'obtiens la loi du logarithme itéré pour une classe de fonctions moins large. En particulier, pour des fonctions dépendant d'un temps d'arrêt, on montre que si T admet un moment à tous les ordres et si F est une fonction de $L^{2+\delta}$, avec $\delta > 0$, alors F vérifie la loi du logarithme itéré. Dans le deuxième paragraphe et pour contourner les difficultés techniques affrontées, j'ai considéré l'algorithme associé à la transformation du shift à droite (voir N. Bouleau [13] où il a étudié le système dynamique $([0, 1]^N, \mathcal{B}([0, 1]^{\otimes N}), \lambda, \theta^{-1})$). Dans cet article [13], est établie la loi du logarithme itéré pour une certaine classe de fonctions dite classe de **Gordin**. Pour ma part, j'obtiens la loi du logarithme itéré et le théorème de la limite centrale pour les fonctions introduites précédemment (chapitre 3 et 4) en vérifiant que ces fonctions sont dans la classe de **Gordin**. Ceci améliore les résultats de [13].

5.1 Sur la loi du logarithme itéré pour l'opérateur de décalage.

Dans la littérature, on trouve des résultats sur le comportement asymptotique de la somme $S_n = \sum_{k=0}^{n-1} \eta_k$, où $(\eta_k)_{k \in \mathbb{N}}$ est considéré comme un processus φ -mélangeant, ou comme une fonctionnelle de processus φ -mélangeant (voir [9],[42]). Plus précisément, on prend

$$\eta_n = f(\zeta_n, \zeta_{n+1}, \dots),$$

où $(\zeta_n)_{n \in \mathbb{N}}$ est un processus φ -mélangeant et f une fonction de $\mathbb{R}^{\mathbb{N}}$ dans \mathbb{R} . La définition d'un processus φ -mélangeant a été rappelée dans le quatrième chapitre. Walter Phillip et William Stout dans [42] ont étudié la somme partielle de plusieurs suites de variables

aléatoires faiblement dépendantes et ils ont proposés des principes d'invariances. En particulier, ils ont traité le cas où $(\eta_n)_{n \in \mathbb{N}}$ est une fonction de processus φ -mélangeant. Il y est établi les résultats rappelés ci-après.

5.1.1 Résultats théoriques existants pour les fonctions pratiquement cylindriques.

Soit $(\zeta_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles. Soit f une application mesurable de l'espace des suites infinies à valeurs réelles $(\alpha_1, \alpha_2, \dots)$. On s'intéresse au processus

$$\eta_n = f(\zeta_n, \zeta_{n+1}, \dots), \quad n \geq 1. \quad (5.1)$$

Ces variables aléatoires sont des fonctionnelles du processus $(\zeta_n)_{n \in \mathbb{N}}$. Walter Phillip et William Stout ont introduit la notion de φ -mélange retardé, condition qui relaxe l'hypothèse de φ -mélange habituelle sur la suite $(\zeta_n)_{n \in \mathbb{N}}$. Si l'on note par \mathcal{F}_a^b la tribu engendrée par les variables aléatoires ζ_n , $a \leq n \leq b$, ils supposent alors que :

$$|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \varphi(nt^{-k}), \quad (5.2)$$

pour tous $A \in \mathcal{F}_1^t$ et pour tous $B \in \mathcal{F}_{t+n}^\infty$, avec k une constante réel positive. On suppose que $\varphi(n)$ est décroissante et que $\lim_{n \rightarrow \infty} \varphi(n) = 0$. Le cas $k = 0$ correspond au cas classique de φ -mélangeance introduit par M. Rosenblatt [47]. Enfin, on suppose qu'il existe pour tout $n \geq 1$ une suite de variables aléatoires $(\eta_{n,l})_{l \in \mathbb{N}}$, \mathcal{F}_n^{n+l} -mesurable, et telle que η_n est approchée dans $L^{2+\delta}$ par la suite $(\eta_{n,l})_{l \in \mathbb{N}}$. On a alors le théorème suivant.

Théorème 14 *Soit $(\zeta_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et f une fonction mesurable de $\mathbb{R}^{\mathbb{N}}$ dans \mathbb{R} , on suppose que $\eta_n = f(\zeta_n, \zeta_{n+1}, \dots)$ vérifie :*

(a)

$$\forall n \geq 1 \quad \mathbb{E}\eta_n = 0. \quad (5.3)$$

Il existe une famille de variables aléatoires réelles $(\eta_{n,l})_{n,l \geq 1}$ et des constantes $0 < \delta \leq 2$ et $C > 0$ telles que

$$\mathbb{E}|\eta_n|^{2+\delta} \leq C \quad (5.4)$$

et, pour tous n et $l \in \mathbb{N}^$*

$$\|\eta_n - \eta_{n,l}\|_{2+\delta} \leq \frac{C}{l^{2+\frac{\delta}{2}}} \quad (5.5)$$

(b) *En plus, on suppose que*

$$\mathbb{E} \left(\sum_{n=1}^N \eta_n \right)^2 = N + O(N^{1-\frac{\delta}{30}}) \quad (5.6)$$

lorsque N tend vers l'infini.

(c) *Enfin, on suppose que ζ_n vérifie la condition de φ -mélange retardé définie par 5.2 avec :*

$$k = \frac{\delta}{11 + 4\delta} \quad (5.7)$$

et

$$\varphi(s) = o\left(s^{-168(1+\frac{2}{\delta})}\right). \quad (5.8)$$

Si l'on définit le processus à temps continu par

$$S_t = \sum_{n \leq t} \eta_n, \quad (5.9)$$

alors il existe un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$, un mouvement Brownien standard $(W(t))_{t \geq 0}$ et $(S(t))_{t \geq 0}$ un processus ayant même loi que $(S_t)_{t \geq 0}$ tels que :

$$S(t) - W(t) = o(t^{\frac{1}{2}-\lambda}) \quad p.s \quad (5.10)$$

$\forall \lambda < \frac{\delta}{588}$, lorsque t tend vers l'infini,

Preuve : Voir chapitre 8 dans les travaux de Walter Phillip et William Stout [42]. ■

A partir de 5.10 et les propriétés classique du mouvement brownien, notamment la loi du logarithme itéré, on déduit que le théorème de la limite centrale et la loi du logarithme itéré sont vérifiés pour le processus η_n .

Reprenons notre système dynamique $([0, 1]^{\mathbb{N}}, \mathcal{B}([0, 1]^{\mathbb{N}}), \lambda, \theta)$ et écrivons ces résultats dans le cadre de nos applications. Soit F une fonction de $L^{2+\delta}([0, 1]^{\mathbb{N}})$ avec $\delta > 0$. Supposons qu'il existe une suite de fonctions $(F_k)_{k \in \mathbb{N}}$, \mathcal{F}_1^k -mesurable, qui approche F dans $L^{2+\delta}$. On a alors le théorème suivant où $F \circ \theta^n$ joue le rôle η_n et $F_k \circ \theta^n$ celui de $\eta_{n,k}$.

Théorème 15 "14.bis" *On suppose que F vérifie les propriétés :*

(a.bis)

$$\mathbb{E}F = 0. \quad (5.11)$$

Il existe $0 < \delta \leq 2$ et C telle que :

$$\mathbb{E}F^{2+\delta} \leq C \quad (5.12)$$

et

$$\|F - F_k\|_{2+\delta} \leq \frac{C}{k^{2+\frac{\delta}{2}}} \quad (5.13)$$

pour tous $k \in \mathbb{N}$.

(b.bis) *En plus, on suppose que :*

$$\mathbb{E} \left(\sum_{k=0}^{N-1} F \circ \theta^k \right)^2 = N + O(N^{1-\frac{\delta}{30}}) \quad (5.14)$$

lorsque N tend vers l'infini.

Alors le processus $(F \circ \theta^k)_{k \in \mathbb{N}}$ vérifie le théorème de la limite centrale à savoir que :

$$\frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F \circ \theta^k \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1),$$

où $\mathcal{N}(0;1)$ est la loi normale centrée réduite et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi. En outre, on a la loi du logarithme itéré, c'est à dire

$$\limsup_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^k}{\sqrt{2N \log \log N}} = 1 \quad \lambda - p.s.$$

et

$$\liminf_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^k}{\sqrt{2N \log \log N}} = -1 \quad \lambda - p.s.$$

Preuve : D'une part, $(U_k)_{k \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes, en particulier, elle est φ -mélangeante, avec $\varphi(n) = 0 \quad \forall n \geq 1$. Ainsi, la condition (c) du théorème 14 est trivialement vérifiée. D'autre part, le processus $(F \circ \theta^k)_{k \in \mathbb{N}}$ est stationnaire. La condition (a.bis) est donc suffisante pour vérifier le (a) du théorème 14. On achève la démonstration en appliquant le théorème précédent. ■

Remarques :

1/ Ces résultats s'étendent pour les fonctions non centrées et vérifiant $\sigma > 0$ en appliquant le théorème à $\frac{F - \mathbb{E}F}{\sigma(F)}$.

2/ Lorsque F est cylindrique il suffit de prendre $F \in L^{2+\delta}$, $\delta > 0$, les autres conditions sont trivialement vérifiées.

Ce théorème, appliqué aux fonctions dépendant d'un temps d'arrêt, fournit les résultats énoncés dans le paragraphe suivant.

5.1.2 Formulation dans le cas des fonctions \mathcal{F}_T -mesurables.

Etudions maintenant les classes de fonctions à temps d'arrêt introduites dans le chapitre 3. Considérons un temps d'arrêt T , $(\mathcal{F}_1^n)_{n \in \mathbb{N}}$ -adapté, et une fonctionnelle F , \mathcal{F}_T -mesurable. On a alors le théorème suivant.

Théorème 16 Soit $\delta \in]0, 2]$. Si l'on suppose que le temps d'arrêt T admet un moment d'ordre p , $p > \max(2 + \frac{\delta}{15}, \frac{2(2+\delta)(1+\delta)(2\delta+7)}{\delta^2})$ alors pour toute $F \in L^{2+2\delta}([0, 1]^{\mathbb{N}}, \mathcal{F}_T)$ d'intégrale nulle le processus $(F \circ \theta^k)_{k \in \mathbb{N}}$ vérifie les propriétés suivantes :

(a) Si $\sigma := \sigma(F) > 0$ on a le théorème de la limite centrale :

$$\frac{1}{\sigma\sqrt{N}} \sum_{k=0}^{N-1} F \circ \theta^k \xrightarrow{\mathcal{L}} \mathcal{N}(0;1),$$

où $\mathcal{N}(0;1)$ est la loi normale centrée réduite et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

(b) La loi du logarithme itéré est vérifiée, c'est à dire

$$\limsup_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^k}{\sqrt{2N \log \log N}} = \sigma \quad \lambda - p.s.$$

et

$$\liminf_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^k}{\sqrt{2N \log \log N}} = -\sigma. \quad \lambda - p.s.$$



Preuve : Sans perte de généralité, on peut prendre $\sigma^2(F) = 1$. Vérifions donc les hypothèses (a.bis) et (b.bis) du théorème 15 ci-avant. Dans la démonstration de la proposition 2 du chapitre 3, on montre que :

$$\mathbb{E} \left(\sum_{n=0}^{N-1} F \circ \theta^n \right)^2 = N - 2 \sum_{k=1}^N k \langle F \circ \theta^k, F \rangle - 2N \sum_{k=N+1}^{\infty} \langle F \circ \theta^k, F \rangle. \quad (5.15)$$

Par ailleurs, si T est un temps d'arrêt ayant un moment d'ordre $p > 2$ alors pour tous $F \in L^2([0, 1]^N, \mathcal{F}_T)$ on a (voir théorème 5 du chapitre 3) :

$$|\langle F \circ \theta^k, F \rangle| \leq \frac{\int_{[0,1]^N} |F|^2 d\lambda (\mathbb{E} T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}}$$

En remplaçons ce résultat dans l'équation 5.15 on trouve que :

$$\begin{aligned} |\mathbb{E} \left(\sum_{n=0}^{N-1} F \circ \theta^n \right)^2 - N| &<< \sum_{k=1}^N k |\langle F \circ \theta^k, F \rangle| + N \sum_{k=N+1}^{\infty} k |\langle F \circ \theta^k, F \rangle| \\ &<< \sum_{k=1}^N k^{1-\frac{p}{2}} + N \sum_{k=N+1}^{\infty} k^{-\frac{p}{2}} \\ &<< N^{2-\frac{p}{2}} \end{aligned}$$

où $<<$ signifie que le terme de gauche est majoré à une constante multiplicative près par le terme à droite. En prenant $p = 2 + \frac{6}{15}$, on vérifie l'hypothèse (b.bis) à savoir que :

$$\mathbb{E} \left(\sum_{k=0}^{N-1} F \circ \theta^k \right)^2 = N + O(N^{1-\frac{6}{30}}).$$

Enfin, si l'on prend $F_k = F \cdot 1_{T \leq k}$ et si l'on pose $r = 2(1 + \delta)$ et $s = \frac{2(2 + \delta)(1 + \delta)}{\delta}$, on a par les inégalités de Hölder et Bienaymé-Tchebichev :

$$\|F - F_k\|_{2+\delta} \leq \|F\|_r \mathbb{P}(T > k)^{\frac{1}{s}} \leq \|F\|_r \frac{\mathbb{E} \left(T^{(2+\frac{r}{\delta})s} \right)^{\frac{1}{s}}}{k^{2+\frac{r}{\delta}}},$$

(5.13) est donc vérifiée dès que T admet un moment d'ordre $\frac{2(2 + \delta)(1 + \delta)(2\delta + 7)}{\delta^2}$. ■

Corollaire 5 *Si T admet un moment à tous les ordres et si F est une fonction de $L^{2+\delta}$, avec $\delta > 0$, alors F vérifie la loi du logarithme itéré et le théorème de la limite centrale.* ■

Lorsque l'on cherche à démontrer, directement, la loi du logarithme itéré pour le processus $(F \circ \theta^k)_{k \in \mathbb{N}}$, par les méthodes utilisées dans les travaux de W. Phillip et W. Stout, on affronte plusieurs problèmes techniques que je ne sais pas résoudre actuellement. En revanche, l'étude de la vitesse de convergence pour la transformation de décalage à droite semble plus adaptée aux résultats théoriques existants (voir [13]). En effet, les premiers résultats, en terme d'opérateur unitaire sur un espace de Hilbert, sont dus à Gordin [19]. Ces résultats ont été développés par la suite par D. J. Scott et C. C. Heyde (voir [27] [49] [23]) et repris dans le cas de variables aléatoires à valeur dans un Banach par Berger (voir [7]). Dans ces travaux, on considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ muni d'une transformation ergodique et inversible τ et on s'intéresse, en particulier, à la vitesse de convergence de la somme $\frac{1}{N} \sum_{k=1}^{N-1} X(\tau^k(\omega))$, où X est une variable aléatoire de carré intégrable et mesurable par rapport à une sous-tribu $\mathcal{M} \subset \mathcal{A}$ vérifiant $\mathcal{M} \subset \tau^{-1}(\mathcal{M})$.

5.2 Sur la loi du logarithme itéré pour l'opérateur de décalage à droite

Considérons $[0, 1]^{\mathbb{Z}}$, l'espace produit de l'intervalle $[0, 1]$, muni de la mesure de Lebesgue produit $\lambda = dx^{\otimes \mathbb{Z}}$ et notons par θ la fonction de décalage, appelée aussi opérateur de shift, définie sur $[0, 1]^{\mathbb{Z}}$ par

$$\theta(\dots, U_{-1}, U_0, U_1, \dots) = (\dots, U_0, U_1, U_2, \dots) \quad \text{i.e.} \quad \theta(U)_k = U_{k+1}$$

Elle est inversible sur $[0, 1]^{\mathbb{Z}}$ et son inverse θ^{-1} est définie par

$$\theta^{-1}(\dots, U_{-1}, U_0, U_1, \dots) = (\dots, U_{-2}, U_{-1}, U_0, \dots).$$

Il est clair que $\theta^{-1}(\lambda) = \lambda$, où $\theta^{-1}(\lambda)$ est la mesure image. Comme dans le cas du shift sur $[0, 1]^{\mathbb{N}}$, on vérifie que la transformation θ^{-1} est fortement mélangeante (on renvoie le lecteur au chapitre 2). Il en résulte donc que le système dynamique $([0, 1]^{\mathbb{Z}}, \mathcal{B}([0, 1]^{\mathbb{Z}}), \lambda, \theta^{-1})$ est ergodique et par conséquent

$$\frac{1}{N} \sum_{k=0}^{N-1} F \circ \theta^{-k} \longrightarrow \mathbb{E}F \quad \lambda - p.s., \quad (5.16)$$

Rappelons que l'intérêt de l'utilisation du théorème ergodique ponctuel de Birkhoff, pour certaines transformations, réside dans son économie de temps de calcul. Il est clair que l'implémentation informatique de l'algorithme déduit (5.16) présente les mêmes particularités que celle du shift. On obtient à nouveau dans ce cadre la loi du logarithme itéré pour certaines classes de fonctions, notamment pour les fonctions F de carré intégrable et mesurables par rapport à la tribu \mathcal{F}_T des événements antérieurs à un temps d'arrêt T . Rappelons que ces fonctions sont très employées en simulation et qu'elles ont fait l'objet de

plusieurs paragraphes dans les chapitres précédents. Dans la suite, on étendra les résultats de N. Bouleau (voir [13]) dans le cas des fonctions dépendants d'un temps d'arrêt, en relaxant les conditions portées sur la fonction F et le temps d'arrêt T . On notera par \mathcal{F}_m^n la tribu engendrée par les coordonnées U_m, \dots, U_n :

$$\mathcal{F}_m^n = \sigma(U_i, m \leq i \leq n) \quad \forall m \leq n \in \mathbb{Z}$$

De la même façon, on notera par :

$$\begin{aligned} \mathcal{F}_{-\infty}^n &= \sigma(U_i, i \leq n) \quad \forall n \in \mathbb{Z} \\ \mathcal{F}_m^{+\infty} &= \sigma(U_i, i \geq m) \quad \forall m \in \mathbb{Z} \\ \mathcal{F}_{-\infty}^{+\infty} &= \sigma(U_i, i \in \mathbb{Z}) \end{aligned}$$

On considère sur $L^1([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^{\infty}, \lambda)$ l'opérateur P défini par :

$$P F(U) = P F(U_1, U_2, \dots) = \int_0^1 F(x, U_1, U_2, \dots) dx \quad \forall F \in L^1(\mathcal{F}_1^{\infty}, \lambda).$$

Il est clair que $P F = \mathbb{E}(F/\mathcal{F}_2^{\infty}) \circ \theta^{-1} \in \mathcal{F}_1^{\infty}$. Par ailleurs, si $(E, || ||)$ est un espace de Banach et P est une contraction de E dans E alors on montre (cf. par exemple [34], [39], [13]) que l'ensemble $(I - P)(E)$ est égale à $\left\{ a \in E / \sup_{n \leq 1} ||a + P(a) + \dots + P^n(a)|| < +\infty \right\}$.

L'opérateur P est ici une contraction de l'espace de Banach $L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^{\infty}, \lambda)$ et il vérifie, donc, le résultat suivant.

Lemme 8 *Soit $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^{\infty}, \lambda)$ et d'intégrale nulle. Les propriétés suivantes sont équivalentes :*

$$(a) \sum_{k=0}^N P^k F \text{ est bornée dans } L^2.$$

$$(b) \sum_{k=0}^N P^k F \text{ converge faiblement dans } L^2.$$

$$(c) \sum_{k=0}^N P^k F \text{ converge dans } L^2.$$

$$(d) \text{ Il existe } G \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^{\infty}, \lambda) \text{ telle que } F = G - P G.$$

Preuve : Nous donnons ici la démonstration proposée par N. Bouleau [13].

$$(b) \Rightarrow (d) : \text{ Si } \sum_{n=0}^N P^n F \text{ converge faiblement alors sa limite } G \text{ appartient à } L^2(\mathcal{F}_1^{\infty}, \lambda).$$

Ceci est dû au théorème de Banach-Steinhaus. La limite G vérifie l'équation de Poisson $F = (I - P) G$.

(d) \Rightarrow (c) : Si l'on considère $F = (I - P) G$ avec $G \in L^2$ d'espérance nulle alors $P^N G$ converge vers 0 lorsque N tend vers l'infini. En effet, $\|P^N G\|_2 = \mathbb{E} \left(\mathbb{E} (G / \mathcal{F}_{N+1}^\infty)^2 \right)$ et $\mathbb{E} (G / \mathcal{F}_{N+1}^\infty)$ est une martingale inverse qui converge vers 0 dans L^2 . Par suite $\sum_{k=0}^N P^k F = G - P^{N+1} G$ converge vers G dans L^2 .

(a) \Rightarrow (b) : Comme $\sum_{k=0}^N P^k F$ est un compact pour la topologie de la convergence faible, il existe une sous-suite $\sum_{k=0}^{N_k} P^k F$ qui converge faiblement dans $L^2(\mathcal{F}_1^\infty, \lambda)$ lorsque k tend vers l'infini. Notons G sa limite. Il résulte de la définition de G que :

$$G - F + \lim_{k \rightarrow \infty} P^{N_k+1} G = P G$$

par les mêmes arguments que l'implication (d) \Rightarrow (c) on montre que $\lim_{k \rightarrow \infty} P^{N_k+1} G = 0$ dans L^2 .

■

On utilisera les mêmes définitions que N. Bouleau (voir [13]) en introduisant la notion de classe de **Gordin**.

Définition 6 On dira qu'une fonction F de $L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, \lambda)$ est dans la classe de **Gordin** et notera $F \in \mathcal{G}$ si $F - \mathbb{E}F$ vérifie l'une des propriétés du lemme 8 ci-avant.

Les fonctions de cette classe vérifient le théorème de la limite centrale et la loi du logarithme itéré. Les démonstrations sont basées sur la décomposition suivante [13].

Proposition 8 Une fonction $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, \lambda)$ est dans la classe de **Gordin** si et seulement si il existe \tilde{G} et $h \in L^2(\mathcal{F}_2^\infty)$ vérifiant $\mathbb{E}(\tilde{G} / \mathcal{F}_2^\infty) = 0$ et $\mathbb{E}h = 0$ telles que

$$F - \mathbb{E}F = \tilde{G} + h \circ \theta - h, \quad (5.17)$$

cette décomposition est unique.

Preuve : Soit une fonction F appartenant à la classe de **Gordin**. D'après le lemme précédent, il existe $G \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, \lambda)$ telle que $F - \mathbb{E}F = G - P G$. Posons :

$$\tilde{G} = G - \mathbb{E}(G / \mathcal{F}_2^\infty)$$

et $h = P G = \mathbb{E}(G / \mathcal{F}_2^\infty) \circ \theta^{-1}$. On obtient ainsi la décomposition voulue. Réciproquement, si F admet une telle décomposition on a alors

$$P \tilde{G} = \mathbb{E}(\tilde{G} / \mathcal{F}_2^\infty) \circ \theta^{-1} = 0$$

et $P(h \circ \theta) = \mathbb{E}(h \circ \theta / \mathcal{F}_2^\infty) \circ \theta^{-1} = h$. Par suite, $F - \mathbb{E}F = (I - P)G$ avec $G = \tilde{G} + h \circ \theta$. Pour l'unicité de la décomposition, prenons $\tilde{G} + h \circ \theta - h = 0$, on déduit que $Ph = h$

puisque $P\tilde{G} = 0$ et $P(h \circ \theta) = h$. Par itération, nous obtenons $h = \lim_{n \rightarrow \infty} P^n h = \mathbb{E}h = 0$ et $\tilde{G} = 0$. ■

En utilisant les résultats de Heyde et Scott (voir [27] [49]), on obtient le théorème de la limite centrale et la loi du logarithme itéré pour l'opérateur de décalage à droite.

Théorème 17 (*N. Bouleau [13]*) *Soit $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, \lambda)$ et d'intégrale nulle. Si l'on suppose que F est dans la classe de Gordin et l'on note par \tilde{G} et h les éléments associés à sa décomposition alors les propriétés suivantes sont vérifiées.*

(a)

$$\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\| \sum_{k=0}^{N-1} F \circ \theta^{-k} \right\|_2 = \|\tilde{G}\|_2,$$

où $\|\cdot\|_2$ désigne la norme L^2 .

(b) Si $\|\tilde{G}\|_2 > 0$ on a le théorème de la limite centrale :

$$\frac{1}{\|\tilde{G}\|_2 \sqrt{N}} \sum_{k=0}^{N-1} F \circ \theta^{-k} \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1),$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

(c) La loi du logarithme itéré est vérifiée, c'est à dire

$$\limsup_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^{-k}}{\sqrt{2N \log \log N}} = \|\tilde{G}\|_2 \quad \lambda - p.s.$$

et

$$\liminf_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^{-k}}{\sqrt{2N \log \log N}} = -\|\tilde{G}\|_2. \quad \lambda - p.s.$$

Preuve : Reprenons la démonstration de N. Bouleau [13]. D'après la décomposition 5.17 on a :

$$\sum_{k=0}^{N-1} F \circ \theta^{-k} = \sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k} + h \circ \theta - h \circ \theta^{-(N-1)}.$$

Il suffit de démontrer que $\lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\| \sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k} \right\|_2 = \|\tilde{G}\|_2$. En effet, par l'inégalité triangulaire on a :

$$\left| \left\| \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} F \circ \theta^{-k} \right\|_2 - \left\| \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k} \right\|_2 \right| \leq 2 \frac{\|h\|_2}{\sqrt{N}}.$$

Ce terme converge vers 0 lorsque N tend vers l'infini. Comme $\mathbb{E}(\tilde{G}/\mathcal{F}_2^\infty) = 0$, on déduit que qu si $k' > k$ $\mathbb{E}(\tilde{G} \circ \theta^{-k} \tilde{G} \circ \theta^{-k'}) = \mathbb{E}(\tilde{G} \circ \theta^{k'-k} \tilde{G}) = 0$ car $\tilde{G} \circ \theta^{k'-k}$ est $\mathcal{F}_{k'-k+1}^\infty \subset \mathcal{F}_2^\infty$ -mesurable. Donc :

$$\| \sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k} \|_2^2 = N \|\tilde{G}\|_2^2.$$

Ceci achève la démonstration de (a). Pour le théorème de la limite centrale, il suffit de démontrer que :

$$\frac{1}{\|\tilde{G}\|_2 \sqrt{N}} \sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k} \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1).$$

Ceci est un résultat de Scott (voir [49]). Pour démontrer (c), on distinguera deux cas. Si $\tilde{G} = 0$, on applique le théorème de Birkhoff à la fonction h^2 et on déduit via le lemme de Kronecker que $\lim_{N \rightarrow \infty} \frac{h \circ \theta^{-N}}{\sqrt{N}} = 0$ λ -p.s. D'où le résultat puisque $\frac{1}{\sqrt{N}} \sum_{k=1}^{N-1} F \circ \theta^{-k} = \frac{h - h \circ \theta^{-N}}{\sqrt{N}}$. Maintenant, si $\tilde{G} \neq 0$, on montre que :

$$\limsup_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k}}{\sqrt{2N \log \log N}} = \|\tilde{G}\|_2 \quad \lambda - p.s.$$

et

$$\liminf_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} \tilde{G} \circ \theta^{-k}}{\sqrt{2N \log \log N}} = -\|\tilde{G}\|_2. \quad \lambda - p.s.$$

Ceci est dû à Heyde et Scott (voir [27]). ■

Remarque : Ces résultats ont été raffinés par Heyde ([26]), où il démontre le théorème de la limite centrale et la loi du logarithme itéré sous leur forme fonctionnelle. Pour plus de détails, on renvoie le lecteur au livre de Hall et Heyde ([23], chapitre 5).

Dans la proposition suivante on fait le lien entre la classe de **Gordin** et la quantité $\sigma^2(F) = \text{Var}(F) + 2 \sum_{k=1}^{\infty} \text{Cov}(F \circ \theta^k, F)$ introduite au chapitre 3.

5.2.1 Classe de Gordin et finitude de σ^2 .

Proposition 9 Soit $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, dx^{\otimes \mathbb{Z}})$ et d'intégrale nulle. Si l'on suppose que F est dans la classe de **Gordin** et l'on note par \tilde{G} et h les éléments associés à sa décomposition alors la série $\sigma^2(F)$ est convergente et

$$\sigma^2(F) = \|\tilde{G}\|_2^2 \tag{5.18}$$

Preuve : Soit $F \in \mathcal{G}$ d'intégrale nulle. En constatant, que l'opérateur P pris sur l'espace $L^2([0, 1]^{\mathbb{N}}, dx^{\otimes \mathbb{N}})$ est l'opérateur de Perron-Frobenius associé à la fonction de décalage θ , on

déduit que $\langle F \circ \theta, F \rangle = \langle F, P F \rangle \quad \forall F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, dx^{\otimes \mathbb{Z}})$. Par suite on a pour tout $N \in \mathbb{N}$:

$$\begin{aligned} \sum_{k=1}^N \langle F \circ \theta^k, F \rangle &= \sum_{k=1}^N \langle F, P^k F \rangle \\ &= \langle F, \sum_{k=1}^N P^k F \rangle. \end{aligned}$$

La fonction étant dans la classe de **Gordin**, la somme $\sum_{k=1}^N P^k F$ est faiblement convergente dans L^2 et par conséquent $\sigma^2(F)$ est fini. Par ailleurs, on a montré dans la proposition 3 du chapitre 3 que si $F \in L^2([0, 1]^{\mathbb{N}})$, d'intégrale nulle et telle que $\sigma^2(F)$ est fini alors :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \int_{[0, 1]^N} \left| \sum_{n=0}^{N-1} F \circ \theta^n \right|^2 d\lambda = \sigma^2(F).$$

Soit encore, puisque $\int_{[0, 1]^N} \left| \sum_{n=0}^{N-1} F \circ \theta^n \right|^2 d\lambda = \int_{[0, 1]^N} \left| \sum_{n=0}^{N-1} F \circ \theta^{-n} \right|^2 d\lambda$,

$$\sigma^2(F) = \lim_{N \rightarrow \infty} \frac{1}{N} \left\| \sum_{n=0}^{N-1} F \circ \theta^{-n} \right\|_2^2 = \|\tilde{G}\|_2^2.$$

■

Une façon de présenter les résultats du théorème 17 et la proposition 9 est de dire : Si $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, dx^{\otimes \mathbb{Z}})$ est dans la classe de **Gordin** alors :

- (a) $\sigma^2(F)$ est une série convergente.
- (b) Si $\sigma^2 := \sigma^2(F) > 0$, on a le théorème de la limite centrale :

$$\frac{1}{\sigma\sqrt{N}} \sum_{k=0}^{N-1} F \circ \theta^{-k} \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1),$$

où $\mathcal{N}(0; 1)$ est la loi normale centrée réduite et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

- (c) La loi du logarithme itéré est vérifié, c'est à dire

$$\limsup_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^{-k}}{\sqrt{2N \log \log N}} = \sigma \quad \lambda - p.s.$$

et

$$\liminf_{N \rightarrow \infty} \frac{\sum_{k=0}^{N-1} F \circ \theta^{-k}}{\sqrt{2N \log \log N}} = -\sigma. \quad \lambda - p.s.$$

Dans la proposition suivante nous rappelons une condition suffisante pour qu'une fonction F soit dans la classe **Gordin**. Ce résultat se trouve dans l'article de N. Bouleau [13].

Proposition 10 Soit $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^\infty, dx^{\otimes \mathbb{Z}})$ telle que :

$$\sum_{k=0}^{\infty} \|\mathbb{E}(F/\mathcal{F}_{k+1}^\infty) - \mathbb{E}F\|_2 \leq +\infty, \quad (5.19)$$

alors F est dans la classe de **Gordin**. En outre, si l'on note par \tilde{G} et h les éléments associés à sa décomposition alors :

$$\|\tilde{G}\|_2 \leq \sum_{k=0}^{\infty} \|\mathbb{E}(F/\mathcal{F}_{k+1}^\infty) - \mathbb{E}F\|_2$$

Preuve : Par définition de P , on a $P F(U_1, U_2, \dots) = \int_0^1 F(x, U_1, U_2, \dots) dx$. En calculant ces itérés on obtient $P^k F(U_1, U_2, \dots) = \int_{[0,1]^k} F(x_1, \dots, x_k, U_1, U_2, \dots) dx_1 \dots dx_k$. Soit encore $P^k F = \mathbb{E}(F/\mathcal{F}_{k+1}^\infty) \circ \theta^{-k}$. Ainsi, on a :

$$\|P^k(F - \mathbb{E}F)\|_2 = \|\mathbb{E}(F/\mathcal{F}_{k+1}^\infty) - \mathbb{E}F\|_2.$$

On déduit, donc, que la série $\sum_{k=0}^N P^k (F - \mathbb{E}F)$ est normalement convergente et donc F est dans la classe de **Gordin**. Si l'on pose $G = \sum_{k=0}^{\infty} P^k (F - \mathbb{E}F)$, alors $\tilde{G} = G - \mathbb{E}(G/\mathcal{F}_1^\infty)$ et on a donc :

$$\|\tilde{G}\|_2 \leq \|G\|_2 \leq \sum_{k=0}^{\infty} \|\mathbb{E}F - \mathbb{E}(F/\mathcal{F}_{k+1}^\infty)\|_2.$$

■

Etudions maintenant les classes de fonctions introduites dans les chapitres précédents. En montrant qu'elles sont dans la classe de **Gordin**, on déduira qu'elles vérifient la loi du logarithme itéré et le théorème de la limite centrale. En particulier, ceci améliore le résultat de N. Bouleau sur les fonctions dépendant d'un temps d'arrêt. Commençons, donc, par les fonctions mesurables par rapport à une tribu \mathcal{F}_T des événements antérieurs à un temps d'arrêt T .

5.2.2 Vitesse de convergence des fonctions \mathcal{F}_T mesurables pour le shift à droite.

Nous travaillons toujours sur le système dynamique $([0, 1]^{\mathbb{Z}}, \mathcal{B}([0, 1]^{\mathbb{Z}}), dx^{\otimes \mathbb{Z}}, \theta^{-1})$. Cependant, nous considérons des temps d'arrêt $(\mathcal{F}_n^\infty)_{n \in \mathbb{N}}$ adaptés et des fonctionnelles \mathcal{F}_T -mesurables. On a alors les résultats suivants.

Théorème 18 Si T admet un moment d'ordre $p > 2$ alors pour tous $F \in L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_T)$, telle que $\int_{[0,1]^{\mathbb{N}}} F = 0$, F est dans la classe de **Gordin**.

Preuve : Montrons que la série (5.19) de la proposition 10 est convergente, c'est à dire

$$\sum_{k=0}^{\infty} \|\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\|_2 \leq +\infty. \text{ Or}$$

$$\begin{aligned} \mathbb{E}\left(\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})^2\right) &= \mathbb{E}\left(F \mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\right) \\ &= \mathbb{E}\left(F \cdot 1_{\{T \leq k\}} \mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\right) + \mathbb{E}\left(F \cdot 1_{\{T > k\}} \mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\right) \end{aligned}$$

En remarquant que $\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})$ est $\mathcal{F}_{k+1}^{\infty}$ -mesurable et que $F \cdot 1_{\{T \leq k\}}$ est \mathcal{F}_0^k -mesurable, on déduit que $\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})$ et $F \cdot 1_{\{T \leq k\}}$ sont indépendantes. D'où

$$\mathbb{E}\left(F \cdot 1_{\{T \leq k\}} \mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\right) = \mathbb{E}\left(F \cdot 1_{\{T \leq k\}}\right) \mathbb{E}\left(\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\right) = 0,$$

Par suite,

$$\begin{aligned} \mathbb{E}\left(\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})^2\right) &= \mathbb{E}\left(F \cdot 1_{\{T > k\}} \mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\right) \\ &= \mathbb{E}\left(F \cdot 1_{\{T > k\}} \mathbb{E}(F/\mathcal{F}_{k+1}^{\infty}) \cdot 1_{\{T > k\}}\right) \\ &\leq \|F \cdot 1_{\{T > k\}}\|_2 \|\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty}) \cdot 1_{\{T > k\}}\|_2. \end{aligned} \quad (5.20)$$

En outre, en constatant que $\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})$ est $\mathcal{F}_{k+1}^{\infty}$ -mesurable, et que $\{T > k\}$ appartient à \mathcal{F}_1^k , on déduit que $\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})$ et $1_{\{T > k\}}$ sont indépendantes. Il ressort que :

$$\mathbb{E}\left(\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})^2\right) \leq \|F \cdot 1_{\{T > k\}}\|_2 \mathbb{P}(T > k)^{\frac{1}{2}} \|\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\|_2.$$

Soit encore :

$$\begin{aligned} \|\mathbb{E}(F/\mathcal{F}_{k+1}^{\infty})\|_2 &\leq \|F \cdot 1_{\{T > k\}}\|_2 \mathbb{P}(T > k)^{\frac{1}{2}} \\ &\leq \|F\|_2 \frac{(\mathbb{E}T^p)^{\frac{1}{2}}}{k^{\frac{p}{2}}}, \end{aligned}$$

puisque $T \in L^p$. ■

Remarque :

1/D'après la démonstration, on observe que la condition (toujours la même voir chapitre 3 et chapitre 4)

$$\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{2}} \cdot \left(\int_{[0,1]^{\mathbb{N}}} |F \cdot 1_{\{T > k\}}|^2 d\lambda\right)^{\frac{1}{2}} < \infty, \quad (5.21)$$

est en fait suffisante pour que F soit dans la classe de **Gordin**.

Prenons, maintenant, la classe des fonctions, F \mathcal{F}_T -mesurables ayant un moment $q \geq 2$, et T ayant un moment d'ordre p , avec $\frac{1}{p} + \frac{1}{q} = 1$. Montrons que ces fonctionnelles sont également dans la classe de **Gordin**.

Théorème 19 Pour toute $F \in L^q([0, 1]^{\mathbb{Z}}, \mathcal{F}_T)$, $q \geq 2$, telle que $\int_{[0, 1]^{\mathbb{N}}} F d\lambda = 0$ et tel que T vérifie $\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{p}} < \infty$ avec $\frac{1}{p} + \frac{1}{q} = 1$, F est dans la classe de Gordin.

Preuve : En reprenant la démonstration du théorème précédent et en appliquant l'inégalité de Hölder au niveau de l'équation (5.20) en lieu et place de l'inégalité de Cauchy-Schwartz, on obtient le résultat suivant.

$$\mathbb{E} \left(\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right)^2 \right) \leq \|F \cdot 1_{\{T > k\}}\|_{L^q} \|\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right) \cdot 1_{\{T > k\}}\|_{L^p}. \quad (5.22)$$

En constatant que $\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right)$ et $1_{\{T > k\}}$ sont indépendants et $p \leq 2$, on déduit :

$$\begin{aligned} \mathbb{E} \left(\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right)^2 \right) &\leq \|F \cdot 1_{\{T > k\}}\|_{L^q} \|\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right)\|_{L^p} \cdot \mathbb{P}(T > k)^{\frac{1}{p}} \\ &\leq \|F \cdot 1_{\{T > k\}}\|_{L^q} \|\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right)\|_2 \cdot \mathbb{P}(T > k)^{\frac{1}{p}}. \end{aligned}$$

Soit encore

$$\|\mathbb{E} \left(F / \mathcal{F}_{k+1}^{\infty} \right)\|_2 \leq \|F \cdot 1_{\{T > k\}}\|_{L^q} \cdot \mathbb{P}(T > k)^{\frac{1}{p}}. \quad (5.23)$$

Ceci achève la démonstration. ■

Remarque : Rappelons que, par application de l'inégalité de Bienaymé-Tchebichev, la condition $\sum_{k=1}^{\infty} \mathbb{P}(T > k)^{\frac{1}{p}} < \infty$ est vérifiée, dès que T admet un moment strictement supérieur à p , et que, lorsque F est bornée, il suffit de supposer T intégrable.

Il reste maintenant à étudier le cas général où la fonction n'est pas nécessairement \mathcal{F}_T -mesurable mais en revanche pratiquement cylindrique. Ces fonctions ont été introduites dans les chapitres précédents.

5.2.3 Vitesse de convergence des fonctions pratiquement cylindriques pour le shift à droite.

Dans cette section, on montre, de la même façon, que si l'on arrive à approcher F , à une vitesse suffisante par une suite de fonctions $(F_k)_{k \in \mathbb{N}}$ ne dépendant que d'un nombre fini de coordonnées, alors F est dans la classe de Gordin et par suite elle vérifie le théorème de la limite centrale et la loi du logarithme itéré. En particulier, on remarquera que les fonctions cylindriques vérifient ces propriétés.

Théorème 20 Soit F une fonctionnelle d'intégrale nulle et ayant un moment d'ordre 2. Si l'on suppose qu'il existe $(F_k)_{k \in \mathbb{N}}$ une suite de $L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^{\infty}, dx^{\otimes \mathbb{Z}}, \theta)$, telle que F_k est \mathcal{F}_1^k -mesurable et que F_k converge vers F dans $L^2([0, 1]^{\mathbb{Z}}, \mathcal{F}_1^{\infty}, \lambda, \theta)$ avec $\sum_{k=1}^{\infty} \|F - F_k\|_2 < \infty$ alors F est dans la classe de Gordin.

Preuve : De la même façon, on écrit :

$$\begin{aligned}\mathbb{E} \left(\mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right)^2 \right) &= \mathbb{E} \left(F \mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right) \right) \\ &= \mathbb{E} \left(F_k \mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right) \right) + \mathbb{E} \left((F - F_k) \mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right) \right)\end{aligned}$$

Le premier terme à droite de l'égalité est nulle puisque F_k et $\mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right)$ sont indépendants. En utilisant l'inégalité de Cauchy-Schwartz, on majore le deuxième terme et on obtient

$$\mathbb{E} \left(\mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right)^2 \right) \leq \|F - F_k\|_2 \|\mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right)\|_2.$$

Soit encore :

$$\|\mathbb{E} \left(F / \mathcal{F}_{k+1}^\infty \right)\|_2 \leq \|F - F_k\|_2. \quad (5.24)$$

■

Corollaire 6 Soit F une fonctionnelle d'intégrale nulle et ayant un moment d'ordre 2. Si l'on suppose que F est \mathcal{F}_1^n -mesurable alors F est dans la classe de Gordin.

■

Application : La méthode de shift est donc aussi efficace pour une intégration en dimension grande, c'est à dire pour l'intégration des fonctions cylindriques.

Remarque : La méthode du shift a une vitesse de convergence du même ordre que la méthode de Monte Carlo. Cependant, on peut dire que sa convergence est contrôlée par le terme,

$$\sigma^2(F) = \text{Var}(F) + 2 \sum_{k=1}^{\infty} \text{Cov}(F \circ \theta^k, F),$$

tandis que la méthode de Monte Carlo, elle, est contrôlée par la variance de F . Or ce terme $\sigma^2(F)$ peut être plus petit que la variance. Dans une telle situation, on aura une convergence théorique plus rapide, sans oublier l'économie du temps de calcul. Auquel cas l'intérêt de la méthode est double. Malheureusement, on ne connaît pas de classe de fonctions raisonnables pour lesquelles on sache démontrer que :

$$\sum_{k=1}^{\infty} \text{Cov}(F \circ \theta^k, F) < 0,$$

même si l'on sait exhiber quelques cas d'école.

Exemples :

1/ Si l'on prend $F - \mathbb{E}F = h \circ \theta - h$ avec $h \in L^2$, on a alors $\sigma^2(F) = 0$. La convergence a lieu en $O(\frac{1}{n})$. On peut construire d'autres exemples à partir de la décomposition de Gordin.

2/ Soit f une fonction de carré intégrable sur $[0, 1]$ et d'intégrale nulle. Si on pose

$$F = a_1 f(U_1) + a_2 f(U_2),$$

alors $\sigma^2(F) = \text{Var}(F) + 2a_1a_2 \int_0^1 f^2(x)dx$. Il suffit donc de prendre a_1 et a_2 de signes contraires.

Ces résultats sont évidemment très artificiels et le problème de trouver des fonctions, mêmes cylindriques, vérifiant $\sigma^2(F) < \text{Var}(F)$, reste ouvert. Il est aussi intéressant de trouver une estimation à priori (théorique ou numérique) du terme σ^2 . Il convient de mentionner que dans le cas des fonctions \mathcal{F}_T -mesurable nous avons, suivant les hypothèses sur les moments de F et T , des majorants relativement explicite pour $\sigma^2(F)$ (voir les remarques à la page 35 et 37).

5.3 Conclusion

Tous les résultats établis ci-dessus renforcent les conclusions faites dans les chapitres précédents à savoir que la méthode du décalage pour le calcul de l'espérance de fonctionnelles de processus aléatoires a une vitesse asymptotique mathématique du même ordre de grandeur que la loi des grands nombres, pour une large classe de fonctionnelles parmi les plus couramment rencontrées et que cette vitesse de convergence est contrôlée par le paramètre σ^2 dont le rôle est similaire à celui que joue la variance dans le cas de la méthode de Monte Carlo.

Chapitre 6

Résolution des équations elliptiques par la méthode du shift

Un champ d'applications naturel des méthodes de simulations est la résolution numérique des équations elliptiques et paraboliques en dimension grande. Dans ce cas les méthodes d'analyse numérique peuvent devenir inutilisables à cause de leur temps de calcul et seules des méthodes probabilistes sont envisageables. Nous avons appliqué la méthode du shift à la résolution des équations elliptiques. Sur un exemple, on a observé qu'une utilisation naïve de la méthode peut aboutir à des problèmes. Une estimation asymptotique du terme $\sigma^2(F)$ en fonction du pas de discrétisation h nous permet d'expliquer les résultats numériques obtenus. Nous proposons, par la suite, d'autres transformations ergodiques dont la vitesse de convergence mathématique est "indépendante" du pas de discrétisation h , pour h petit. A la fin de ce chapitre nous suggérons que la méthode du shift soit employée sur d'autres représentations du mouvement brownien.

6.1 Représentation probabiliste des solutions d'une équation aux dérivées partielles

Soit L un opérateur elliptique défini sur un domaine borné $D \in \mathbb{R}^n$ de bord régulier. On considère l'opérateur linéaire du second ordre :

$$L u = \frac{1}{2} \sum_{i,j=1}^n a_{i,j}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u}{\partial x_i} + c(x) u \quad (6.1)$$

où $(a_{i,j})_{1 \leq i,j \leq n}$, $(b_i)_{1 \leq i \leq n}$ et c sont des fonctions du domaine D dans \mathbb{R} . On suppose que l'opérateur L est elliptique c'est à dire que pour tout $x \in D$ la matrice $(a_{i,j}(x))_{1 \leq i,j \leq n}$ est symétrique définie positive. On cherche à résoudre le problème de Dirichlet suivant :

$$\begin{cases} L u(x) = f(x) & \text{dans } D \\ u(x) = \phi(x) & \text{sur } \partial D \end{cases} \quad (6.2)$$

où f et ϕ sont des fonctions régulières de \mathbb{R}^n dans \mathbb{R} . Sous certaines conditions de régularité, le système (6.2) admet une solution unique (voir Friedman [17]). Cette solution admet une représentation en termes de solution d'équation différentielle stochastique. En effet, la matrice

$(a_{ij}(x))$ étant définie positive, il existe alors une unique matrice $(\sigma_{ij}(x))$ définie positive vérifiant, $\sigma(x)\sigma(x) = a(x)$. De plus, on montre que si $(a_{ij}(x))$ est lipschitzienne sur \bar{D} alors $(\sigma_{ij}(x))$ l'est aussi (voir Friedman [17]). On prolonge $\sigma(x)$ et $b(x)$ sur \mathbb{R}^n , telle que

$$|\sigma(x) - \sigma(y)| \leq C|x - y|, \quad |b(x) - b(y)| \leq C|x - y|$$

On considère l'équation différentielle stochastique

$$dx_t = b(x_t)dt + \sigma(x_t)dw_t \quad \text{et} \quad x_0 = x \in D \quad (6.3)$$

où $(w_t)_{t \leq 0}$ est un mouvement Brownien standard sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Théorème 21 *Si l'on suppose que :*

- $(a_{ij})_{1 \leq i, j \leq n}, (b_i)_{1 \leq i \leq n}$ sont lipschitziennes sur l'adhérence \bar{D} de D
- $c \leq 0$, hölderienne sur \bar{D}
- l'opérateur L est uniformément elliptique sur D , c'est-à-dire : il existe $\mu > 0$ tel que

$$\forall x \in D, \forall \xi \in \mathbb{R}^n \quad \sum_{i,j=1}^n a_{ij}(x) \xi_i \xi_j \geq \mu |\xi|^2.$$

- la frontière de D est de classe C^2
- f est une fonction hölderienne sur \bar{D}
- ϕ est continue sur ∂D

alors le problème de Dirichlet (6.2) admet une solution unique et cette solution est donnée par :

$$u(x) = \mathbb{E}_x \left(\phi(x_\tau) \exp \left\{ \int_0^\tau c(x_s) ds \right\} \right) - \mathbb{E}_x \left(\int_0^\tau f(x_t) \exp \left\{ \int_0^t c(x_s) ds \right\} dt \right) \quad (6.4)$$

où $\tau := \inf \{t \geq 0 \text{ tel que } x_t \notin D\}$.

■

La démonstration de ce résultat se trouve dans le livre de Friedman "stochastic differential equations and applications", voir [17], théorème 5.1, page 145.

Considérons le cas particulier où les fonctions f et c sont nulles. Soit, donc, L l'opérateur différentiel défini par

$$L u = \frac{1}{2} \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u}{\partial x_i}, \quad (6.5)$$

on cherche à résoudre le problème de Dirichlet

$$\begin{cases} L u = 0 & \text{dans } D \\ u = \phi & \text{sur } \partial D \end{cases} \quad (6.6)$$

D'après ce qui précède, la solution est donnée par :

$$u(x) = \mathbb{E}_x \phi(x_\tau) \quad (6.7)$$

avec $dx_t = b(x_t)dt + \sigma(x_t)dw_t$ et $\tau = \inf \{t \geq 0 \text{ tel que } x_t \notin D\}$. De même, si l'on prend ϕ et c nulle et si l'on suppose que f est une fonction constante, alors la solution du système :

$$\begin{cases} Lu = C & \text{dans } D \\ u = 0 & \text{sur } \partial D \end{cases} \quad (6.8)$$

est donnée par $u(x) = -C\mathbb{E}_x(\tau)$.

Rappelons que les techniques d'analyse numérique "classique" pour la résolution des équations elliptiques en grande dimension $n \geq 4$ peuvent devenir inutilisables à cause de leur temps de calcul. Par suite, une méthode de type Monte Carlo est envisageable. Par ailleurs, le calcul numérique des solutions des systèmes (6.6) et (6.8) se ramène à des intégrations en dimension infinie, et il se trouve donc très adapté à la méthode du shift. Pour calculer la solution u en un point x donné, nous avons utilisé ce procédé pour approcher $\mathbb{E}_x(\tau)$. Une discrétisation classique par la méthode d'Euler de l'équation différentielle stochastique nous donne l'algorithme suivant

$$\begin{cases} \bar{x}_0^h &= x \\ \bar{x}_{k+1}^h &= \bar{x}_k^h + b(\bar{x}_k^h)h + \sigma(\bar{x}_k^h)\sqrt{h} g_{k+1} \end{cases}$$

avec $(g_i)_{i \in \mathbb{N}}$ est une suite de gaussiennes centrées réduites. Dans un premier temps, nous avons utilisé la méthode d'une façon naïve (des applications moins naïves de la méthode du shift sont indiquées à la fin de ce chapitre). On a calculé τ^h aux points successifs

$$\begin{aligned} g &= (g_1, g_2, \dots, g_k, \dots) \\ \theta(g) &= (g_2, g_3, \dots, g_{k+1}, \dots) \\ &\vdots \end{aligned}$$

et on a pris la moyenne des termes. Nous avons testé ce procédé sur des exemples simple et on a obtenu les résultats numériques suivants.

6.2 Résultats numériques

D'après ce qui précède, la solution du système différentiel

$$\begin{cases} \frac{1}{2}u''(x) &= -1 \quad \forall x \in [-2, 2] \\ u(2) &= u(-2) = 0 \end{cases}$$

est donnée par $u(x) = \mathbb{E}_x(\tau)$, avec τ est le temps de sortie du mouvement Brownien standard $w(t)$ de l'intervalle $[-2, 2]$. Il est clair que $u(x) = 4 - x^2$ sur l'intervalle $[-2, 2]$. Pour calculer u au point 0, on considère l'algorithme

$$\begin{cases} \bar{x}_0^h &= 0 \\ \bar{x}_{k+1}^h &= \bar{x}_k^h + \sqrt{h} g_{k+1} \end{cases}$$

avec $(g_i)_{i \in \mathbb{N}}$ est une suite de gaussiennes centrées réduites et h est le pas de discrétisation. On approche $u(x) = \mathbb{E}_x(\tau)$ par $\mathbb{E}_x(h\tau^h)$, où $h\tau^h$ est le temps de sortie de la chaîne approximante. Une comparaison de la méthode du shift et de la méthode de Monte Carlo avec des pas de discrétisation de plus en plus fins, a donné les résultats suivants :

- La représentation graphique ci-dessous représente la convergence du shift et de Monte Carlo avec un pas de discrétisation $h = 0.1$.

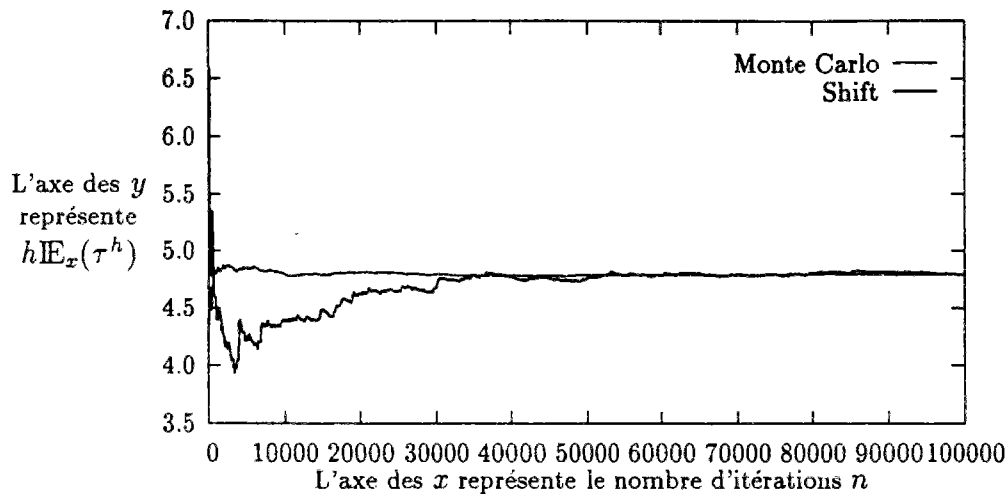


Figure 6.1

- La représentation graphique ci-dessous représente la convergence du shift et de Monte Carlo avec un pas de discrétisation $h = 0.001$.

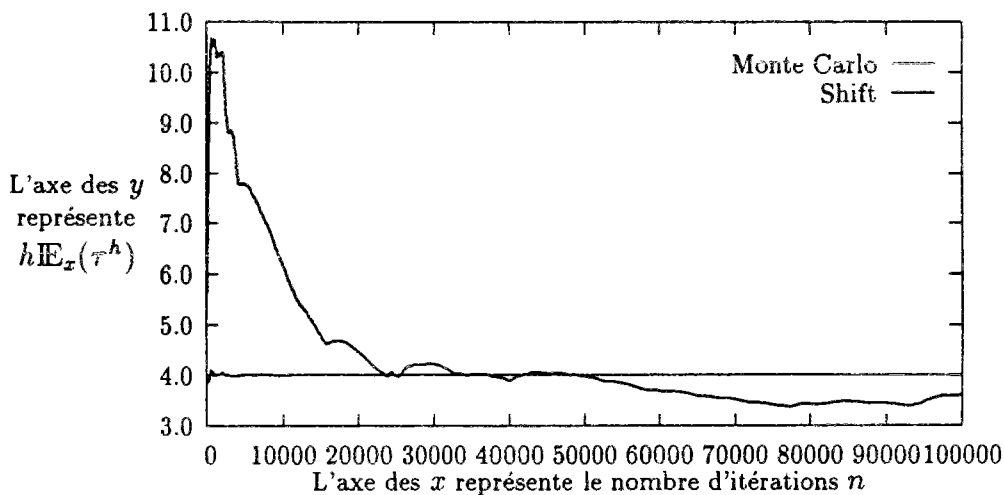


Figure 6.2

Ces courbes permettent d'observer que par la méthode du shift et avec un pas de discrétisation $h = 0.1$, une précision de 10^{-2} sur $h\mathbb{E}_x(\tau^h)$ est définitivement atteinte à partir de 40000 itérations alors que pour une valeur plus fine $h = 0.001$, après 40000 itérations, on observe encore des fluctuations de plus de 5.10^{-1} en valeur absolue. On remarque donc que la vitesse de convergence dépend du pas de discrétisation. Ceci peut être expliqué par le raisonnement heuristique suivant. Pour un pas de discrétisation h donné, on calcule notre fonctionnelle sur une trajectoire brownienne $\omega \in \mathcal{C}([0, +\infty[, \mathbb{R})$, puis sur la trajectoire décalée d'un pas de temps h ; après plusieurs itérations on prend la moyenne de Césaro. Le problème se présente lorsque h est trop petit, auquel cas on reste pratiquement sur la même trajectoire et on augmente, donc, les corrélations. Les justifications théoriques sont données dans le paragraphe suivant.

6.3 Interprétation des résultats obtenus

Désignons par $\Omega = \mathcal{C}([0, +\infty[, \mathbb{R})$, l'espace des fonctions continues sur l'intervalle $[0, +\infty[$. On munit Ω de la mesure de Wiener \mathbb{P} . Les trajectoires du mouvement brownien $(w_t)_{t \geq 0}$ sont définies par : $\forall \omega \in \Omega, w_t(\omega) = \omega(t)$. Pour h fixé on notera θ_h la transformation du shift définie sur Ω par $\theta_h w(\cdot) = w(\cdot + h)$. On désigne par $(\mathcal{F}_t)_{t \geq 0}$ la filtration naturelle complétée du brownien (donc continue à droite puisque $(w_t)_{t \geq 0}$ est un processus à accroissements indépendants stationnaire). Considérons l'équation différentielle stochastique définie par :

$$\begin{cases} x_0 &= x, \quad x \in \mathbb{R} \\ dx_t &= b(x_t)dt + \sigma(x_t)dw_t \end{cases} \quad (6.9)$$

où b et σ sont deux fonctions à valeurs réelles. On cherche à calculer $\mathbb{E}T$ où T est le temps de sortie du processus $(x_t)_{t \geq 0}$, d'un ensemble A donné, $T = \inf\{t \geq 0 \text{ tel que } x_t \notin A\}$. Sous certaines conditions sur A , le temps aléatoire T est un $(\mathcal{F}_t)_{t \geq 0}$ -temps d'arrêt (exemple A est un fermé). Une discrétisation de l'équation (6.9) par le schéma d'Euler nous suggère l'algorithme suivant :

$$\begin{cases} \bar{x}_0^h &= x \\ \bar{x}_{k+1}^h &= \bar{x}_k^h + b(\bar{x}_k^h)h + \sigma(\bar{x}_k^h)\sqrt{h} g_{k+1} \end{cases} \quad (6.10)$$

avec $(g_i)_{i \in \mathbb{N}}$ est une suite de gaussiennes centrées réduites. Si l'on note par τ^h le temps de sortie de la chaîne $(\bar{x}_k^h)_{k \in \mathbb{N}}$,

$$\tau^h = \inf\{k \geq 0, \bar{x}_k^h \notin A\},$$

alors $\mathbb{E}_x(T)$ est approchée par $\mathbb{E}_x(h\tau^h)$. En outre, si l'on pose

$$\bar{x}_t^h = \bar{x}_k^h \quad \forall t \in [kh, (k+1)h[,$$

il est clair que le temps de sortie du processus $(\bar{x}_t^h)_{t \geq 0}$ vérifie $T^h := h\tau^h$. On dira par la suite que T^h est le temps de sortie de la chaîne approximante.

Il résulte des chapitres précédents que la vitesse de convergence pour le calcul de $\mathbb{E}_x(T^h)$ par la méthode du shift est contrôlée par :

$$\sigma^2(T^h, \theta) = \text{Var}(T^h) + 2 \sum_{k=1}^{\infty} \text{Cov}(T^h \circ \theta^k, T^h)$$

où θ est le shift "habituel" défini sur $\mathbb{R}^{\mathbb{N}}$ par $\theta(g_1, g_2, \dots) = (g_2, g_3, \dots)$. Pour interpréter les résultats numériques obtenus, on s'intéressera au comportement asymptotique de $\sigma^2(T^h)$. Par ailleurs, pour définir $(g_k)_{k \in \mathbb{N}}$ sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ que le mouvement brownien $(w_t)_{t \geq 0}$, il est commode de prendre $g_k^h = \frac{w_{kh} - w_{(k-1)h}}{\sqrt{h}}$ pour $k \geq 1$. Il est intéressant de noter que dans ce cas :

$$\begin{aligned} T^h \circ \theta^k(g_1^h(\omega), g_2^h(\omega), \dots) &= T^h(g_{k+1}^h(\omega), g_{k+2}^h(\omega), \dots) \\ &= T^h \circ \theta_{kh}(\omega), \end{aligned}$$

et que par suite $\sigma^2(T^h, \theta) = \text{Var}(T^h) + 2 \sum_{k=1}^{\infty} \text{Cov}(T^h \circ \theta_{kh}, T^h)$. De plus, le temps de sortie T^h de la chaîne approximante est aussi un $(\mathcal{F}_t)_{t \geq 0}$ -temps d'arrêt.

Pour plus de généralité, on considérera donc une variable aléatoire X , \mathcal{F}_T -mesurable et on supposera qu'il existe un processus $(X^h)_{h>0}$ qui approche X dans L^2 lorsque h converge vers 0 et que X^h est \mathcal{F}_{T^h} -mesurable pour tout $h > 0$. Dans la pratique le processus est obtenu à partir du schéma d'Euler. On désigne par

$$\sigma^2(h) := \sigma^2(X^h, \theta) = \text{Var}(X^h) + 2 \sum_{k=1}^{\infty} \text{Cov}(X^h \circ \theta_{kh}, X^h), \quad (6.11)$$

et par

$$\tau^2(h) := \sigma^2(X, \theta) = \text{Var}(X) + 2 \sum_{k=1}^{\infty} \text{Cov}(X \circ \theta_{kh}, X). \quad (6.12)$$

Dans la proposition suivante on montre que $\tau^2(h)$ est asymptote à $\sigma^2(h)$ lorsque h tend vers zéro.

Proposition 11 Soit $\varepsilon > 0$, fixé. Soient T et $(T^h)_{h>0}$ des $(\mathcal{F}_t)_{t \geq 0}$ -temps d'arrêts vérifiant :

$$\begin{cases} T \leq T^h \\ \sup_{h>0} \|T^h\|_{2+\varepsilon} < \infty \end{cases}$$

X est une variable aléatoire \mathcal{F}_T -mesurable et $\forall h > 0$, X^h est \mathcal{F}_{T^h} -mesurable. Si X^h et $X \in L^2$ alors $\sigma^2(h)$ et $\tau^2(h)$ sont finis. En outre,

$$\text{si } \text{Var}(X - X^h) = o(h^{\frac{\varepsilon}{2}}) \text{ alors } \lim_{h \rightarrow 0} h (\sigma^2(h) - \tau^2(h)) = 0. \quad (6.13)$$

Preuve : Pour ce faire, on commencera par remarquer que la fonction, $F \rightarrow \sigma(F) := \sqrt{\sigma^2(F)}$, est une semi-norme sur L^2 . En effet, d'après la proposition 2 du chapitre 3, on a :

$$\sigma^2(F) = \lim_{N \rightarrow \infty} \frac{1}{N} \text{Var}\left(\sum_{n=0}^{N-1} F \circ \theta^n\right).$$

Soit encore

$$\sigma(F) = \lim_{N \rightarrow \infty} \frac{1}{\sqrt{N}} \left\| \sum_{n=0}^{N-1} (F - \mathbb{E}F) \circ \theta^n \right\|_2,$$

et, Par conséquent, $\sigma(F + G) \leq \sigma(F) + \sigma(G)$ et $\sigma(\lambda F) = |\lambda|\sigma(F)$ pour tous $F, G \in L^2$ et $\lambda \in \mathbb{R}$. Il suffit, donc, de montrer que $\lim_{h \rightarrow 0} h\sigma^2(X - X^h) = 0$, puisque :

$$\begin{aligned} |\sigma(h) - \tau(h)| &= |\sigma(X) - \sigma(X^h)| \\ &\leq \sigma(X - X^h). \end{aligned}$$

Or ce terme est par définition égal à

$$\sigma^2(X - X^h) := \text{Var}(X - X^h) + 2 \sum_{k=1}^{\infty} \text{Cov}((X - X^h) \circ \theta_{kh}, (X - X^h)) \quad (6.14)$$

Le premier terme à droite de l'égalité (6.14) converge vers 0 puisque X^h converge vers X dans L^2 . Les termes de la série se décomposent comme suit :

$$\begin{aligned} \text{Cov}((X - X^h) \circ \theta_{kh}, (X - X^h)) &= \text{Cov}((X - X^h) \circ \theta_{kh}, (X - X^h) \cdot 1_{\{T^h \leq kh\}}) \\ &\quad + \text{Cov}((X - X^h) \circ \theta_{kh}, (X - X^h) \cdot 1_{\{T^h > kh\}}). \end{aligned} \quad (6.15)$$

Comme $T \leq T^h$ on déduit que la variable aléatoire X est \mathcal{F}_{T^h} -mesurable et par suite $(X - X^h) \cdot 1_{\{T^h \leq kh\}}$ est \mathcal{F}_{kh} -mesurable. Par conséquent, les variables aléatoires $X^h \circ \theta_{kh}$ et $(X - X^h) \cdot 1_{\{T^h \leq kh\}}$ sont indépendantes. Ceci montre que le premier terme de l'égalité (6.15) est nul. En appliquant Cauchy-Schwartz au deuxième terme, il vient que

$$|\text{Cov}((X - X^h) \circ \theta_{kh}, (X - X^h))| \leq \sqrt{\text{Var}((X - X^h) \circ \theta_{kh} \cdot 1_{\{T^h > kh\}})} \sqrt{\text{Var}(X - X^h)}$$

En constatant que $1_{\{T^h > kh\}}$ est \mathcal{F}_{kh} -mesurable on déduit que :

$$\begin{aligned} |\text{Cov}((X - X^h) \circ \theta_{kh}, (X - X^h))| &\leq \text{Var}(X - X^h) \sqrt{\mathbb{P}(T^h > kh)} \\ &\leq \text{Var}(X - X^h) \frac{\sqrt{\mathbb{E}(T^h)^{2+\varepsilon}}}{k^{1+\frac{\varepsilon}{2}} h^{1+\frac{\varepsilon}{2}}} \\ &<< \frac{\text{Var}(X - X^h)}{k^{1+\frac{\varepsilon}{2}} h^{1+\frac{\varepsilon}{2}}} \end{aligned}$$

où $<<$ signifie que le terme à gauche est majoré à une constante multiplicative près par le terme à droite. A ce stade, en faisant X ou $X^h = 0$ on retrouve comme d'habitude que $\sigma^2(h)$ et $\tau^2(h)$ sont finis. Enfin, en introduisant cette inégalité dans l'équation 6.14 on obtient et en multipliant par h :

$$h\sigma^2(X - X^h) << \frac{\text{Var}(X - X^h)}{h^{\frac{\varepsilon}{2}}}.$$

Ceci achève la démonstration. ■

Remarque : Malheureusement, en parcourant la littérature on est surpris de ne pas trouver des estimations de $\text{Var}(T - T^h)$. Cependant, il nous semble possible qu'il existe $\varepsilon > 0$ tel que $\text{Var}(T - T^h) = o(h^\varepsilon)$.

Montrons maintenant que la somme $\tau^2(h)$ diverge lorsque h converge vers zéro. Pour ce faire on montrera que la fonction qui à t fait correspondre $Cov(X \circ \theta_t, X)$ est continue et qu'elle est intégrable sur $[0, \infty[$. Pour l'intégrabilité on montre (cf. la démonstration ci-dessus) que si T admet un moment d'ordre $2 + \varepsilon$ et $X \in L^2$ alors $Cov(X \circ \theta_t, X) < \frac{1}{t^{1+\frac{\varepsilon}{2}}}$. La continuité est vraie dès que $X \in L^2$.

Lemme 9 Soit $X \in L^2$ alors la fonction $\varphi_X(s) = Cov(X \circ \theta_s, X)$ est continue sur $[0, \infty[$.

Preuve : On remarque que :

$$\begin{aligned} |\varphi_X(s) - \varphi_X(h)| &= |\mathbb{E}(X \circ \theta_s X) - \mathbb{E}(X \circ \theta_h X)| \\ &\leq \|X \circ \theta_{|s-h|} - X\|_2 \|X\|_2. \end{aligned}$$

Il suffit, donc, de montrer $\lim_{\substack{s \rightarrow 0 \\ s > 0}} X \circ \theta_s = X$ dans L^2 . Comme la topologie de la convergence uniforme sur les compacts, définie sur $\mathcal{C}([0, +\infty[, \mathbb{R})$, est métrique séparable, il existe une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ continues bornées définies de $\mathcal{C}([0, +\infty[, \mathbb{R})$ dans \mathbb{R} qui approche X dans L^2 . Dans ce cas :

$$\begin{aligned} \|X \circ \theta_s - X\|_2 &\leq \|X \circ \theta_s - X_n \circ \theta_s\|_2 + \|X - X_n\|_2 + \|X_n \circ \theta_s - X_n\|_2 \\ &\leq 2\|X - X_n\|_2 + \|X_n \circ \theta_s - X_n\|_2, \end{aligned}$$

et par suite

$$\limsup_{s \rightarrow 0} \leq 2\|X - X_n\|_2 \quad \forall n \in \mathbb{N}.$$

Sans perte de généralité, on peut donc supposer X une variable aléatoire continue bornée. Or, si la variable aléatoire X est continue pour la topologie de la convergence uniforme sur les compacts sur $\mathcal{C}([0, +\infty[, \mathbb{R})$ alors :

$$\forall \omega \in \mathcal{C}([0, +\infty[, \mathbb{R}) \quad \lim_{s \rightarrow 0} X \circ \theta_s(\omega) - X(\omega) = 0$$

et si on suppose qu'elle est bornée alors on a convergence dans L^2 . ■

Le lemme suivant nous donne l'intégrabilité de la fonction φ_X sur $[0, \infty[$.

Lemme 10 Soit $\varepsilon > 0$, fixé. Soit T un $(\mathcal{F}_t)_{t \geq 0}$ - temps d'arrêts de $L^{2+\varepsilon}$ et X est une variable aléatoire \mathcal{F}_T -mesurable alors

$$|\varphi_X(s)| \leq Var(T) \frac{\sqrt{\mathbb{E}(T^{2+\varepsilon})}}{s^{1+\frac{\varepsilon}{2}}} \quad (6.16)$$

et par conséquent $\varphi_X \in L^1(\mathbb{R}_+, dx)$.

Preuve : En utilisant le même raisonnement que la proposition précédente, on montre que :

$$\begin{aligned} |Cov(T \circ \theta_s, T)| &= |Cov(T \circ \theta_s, T \cdot 1_{\{T > s\}})| \\ &\leq Var(T) \sqrt{\mathbb{P}(T > s)} \\ &\leq Var(T) \frac{\sqrt{\mathbb{E}(T^{2+\varepsilon})}}{s^{1+\frac{\varepsilon}{2}}}. \end{aligned}$$

Ceci achève la démonstration. ■

Proposition 12 Soit $\varepsilon > 0$, fixé. Soit T un $(\mathcal{F}_t)_{t>0}$ -temps d'arrêts de $L^{2+\varepsilon}$ et X est une variable aléatoire \mathcal{F}_T -mesurable alors $\lim_{h \rightarrow 0} h\tau^2(h) = 2 \int_0^\infty \varphi(s)ds$.

Preuve :

$$\begin{aligned} |h\tau^2(h) - 2 \int_0^\infty \varphi(s)ds| &\leq h\text{Var}(T) + 2|h \sum_{kh \leq A} \varphi(hk) - \int_0^A \varphi(s)ds| \\ &\quad + 2h \sum_{kh > A} |\varphi(hk)| + 2 \int_A^\infty |\varphi(s)|ds. \end{aligned}$$

D'après le lemme 10, $|\varphi(hk)| \ll \frac{1}{k^{1+\frac{\varepsilon}{2}}h^{1+\frac{\varepsilon}{2}}}$, où \ll signifie que le terme à gauche est plus petit que le terme à droite à une constante multiplicative près, on déduit que

$$h \sum_{kh > A} |\varphi(hk)| \ll \int_A^\infty \frac{dx}{x^{1+\frac{\varepsilon}{2}}} \ll \frac{1}{A^{\frac{\varepsilon}{2}}}$$

et, donc, on peut choisir A assez grand pour que le troisième terme à droite de l'inégalité soit aussi petit qu'on veut uniformément en h . Il est de même pour le quatrième terme puisque $\int_A^\infty |\varphi(s)|ds \ll \int_A^\infty \frac{dx}{x^{1+\frac{\varepsilon}{2}}} \ll \frac{1}{A^{\frac{\varepsilon}{2}}}$. Pour A choisi, le deuxième terme, lui, converge vers 0 lorsqu'on fait tendre h vers 0 puisque φ est continue et donc Riemann intégrable (voir lemme 9). Ceci achève la démonstration. ■

Corollaire 7 Soit $\varepsilon > 0$, fixé. Soient T et $(T^h)_{h>0}$ des $(\mathcal{F}_t)_{t>0}$ -temps d'arrêts vérifiant :

$$\begin{cases} T \leq T^h \\ \sup_{h>0} \|T^h\|_{2+\varepsilon} < \infty \end{cases}$$

Si l'on suppose que $\text{Var}(T - T^h) = o(h^{\frac{\varepsilon}{2}})$ alors $\sigma^2(h) = \sigma^2(T^h) = O(\frac{1}{h})$

Preuve : Ce résultat découle des deux propositions précédentes. ■

Remarque : D'après le corollaire ci-dessus et les résultats numériques obtenus on conjecture qu'il existe ε tel que $\text{Var}(T - T^h) = o(h^\varepsilon)$ et par conséquent $\sigma^2(h) = O(\frac{1}{h})$.

Cette étude asymptotique de $\sigma^2(h)$, pour h petit, nous a permis de choisir d'autres transformations ergodiques dont la vitesse de convergence mathématique est "indépendante" du pas de discrétisation, pour h petit. En outre, pour ces transformations on a conservé les particularités de l'implémentation informatique de la méthode du shift, c'est à dire une vitesse de convergence informatique plus grande que la méthode de Monte Carlo.

6.4 Application de la méthode du shift à la résolution des équations elliptiques.

Pour diminuer les corrélations, nous avons décalé les trajectoires d'un pas de temps indépendant du pas de discrétisation. Autrement dit, pour un pas de discrétisation h donné,

on calcule notre fonctionnelle sur une trajectoire brownienne $\omega \in \mathcal{C}([0, +\infty[, \mathbb{R})$, puis sur la trajectoire décalée d'un pas de temps α , $\alpha > 0$; après plusieurs itérations on prend la moyenne de Césaro. Pour simplifier les notations, on prendra $\alpha = 1$ et $h = \frac{1}{n}$, $n \in \mathbb{N}$. Rappelons que l'espérance du temps d'arrêt T est approchée par l'espérance du temps de sortie de la chaîne approximante T^h et que $T^h = h\tau^h$ où τ^h est le temps de sortie du schéma discrétisé. On a aussi considéré la discrétisation de l'équation (6.9) donnée par l'algorithme :

$$\begin{cases} \bar{x}_0^h &= x \\ \bar{x}_{k+1}^h &= \bar{x}_k^h + b(\bar{x}_k^h)h + \sigma(\bar{x}_k^h)\sqrt{h} g_{k+1}^h \end{cases}$$

avec $g_k^h = \frac{w_{kh} - w_{(k-1)h}}{\sqrt{h}}$. Le procédé proposé au début du paragraphe consiste à calculer τ^h aux points successifs $g^h, \theta^n(g^h), \theta^{2n}(g^h), \dots$ avec

$$\begin{aligned} g^h &= (g_1^h, g_2^h, \dots, g_n^h, \dots, g_k^h, \dots) \\ \theta^n(g^h) &= (g_{n+1}^h, g_{n+2}^h, \dots, g_{2n}^h, \dots, g_{n+k}^h, \dots) \\ &\vdots \end{aligned}$$

où θ est le shift "habituel" sur \mathbb{R}^N . D'après les résultats des chapitres précédents, la vitesse de convergence, pour l'opérateur θ^n , est contrôlée par :

$$\begin{aligned} \sigma^2(T^h, \theta^n) &= \text{Var}(T^h) + 2 \sum_{k=1}^{\infty} \text{Cov}(T^h \circ \theta^{nk}, T^h) \\ &= \text{Var}(T^h) + 2 \sum_{k=1}^{\infty} \text{Cov}(T^h \circ \theta_{nkh}, T^h). \end{aligned}$$

En particulier et parce qu'on a choisit $nh = 1$ on a alors :

$$\sigma^2(T^h, \theta^n) = \text{Var}(T^h) + 2 \sum_{k=1}^{\infty} \text{Cov}(T^h \circ \theta_k, T^h).$$

On cherche à montrer que $\lim_{h \rightarrow 0} \sigma^2(T^h, \theta^n) = \sigma^2(T, \theta^n) = \text{Var}(T) + 2 \sum_{k=1}^{\infty} \text{Cov}(T \circ \theta_k, T)$.

Reprenons le cas général où X est \mathcal{F}_T -mesurable et X^h est \mathcal{F}_{T^h} -mesurable pour tout $h > 0$. Si l'on note par $\tilde{\sigma}^2(h) := \sigma^2(X^h, \theta^n)$ et par $\tilde{\tau}^2 := \sigma^2(X, \theta^n)$ on a alors le résultat suivant.

Proposition 13 Soit $\varepsilon > 0$, fixé. Soient T et $(T^h)_{h>0}$ des $(\mathcal{F}_t)_{t>0}$ -temps d'arrêts vérifiant :

$$\begin{cases} T \leq T^h \\ \sup_{h>0} \|T^h\|_{2+\varepsilon} < \infty \end{cases}$$

X est une variable aléatoire \mathcal{F}_T -mesurable et $\forall h > 0$, X^h est \mathcal{F}_{T^h} -mesurable. Si X^h et $X \in L^2$ alors $\tilde{\sigma}^2(h)$ et $\tilde{\tau}^2$ sont finis. En outre,

$$\lim_{h \rightarrow 0} \tilde{\sigma}^2(h) = \tilde{\tau}^2. \quad (6.17)$$

Preuve : Par la même démonstration que la proposition 11, on a

$$\begin{aligned} |\tilde{\sigma}(h) - \tilde{\tau}| &= |\sigma(X, \theta^n) - \sigma(X^h, \theta^n)| \\ &\leq \sigma((X - X^h), \theta^n). \end{aligned}$$

Il suffit, donc, de montrer que $\lim_{h \rightarrow 0} \tilde{\sigma}^2(X - X^h) = 0$ où

$$\tilde{\sigma}^2(X - X^h) := \text{Var}(X - X^h) + 2 \sum_{k=1}^{\infty} \text{Cov}((X - X^h) \circ \theta_k, (X - X^h)) \quad (6.18)$$

Le premier terme à droite de l'égalité (6.18) converge vers 0 puisque X^h converge vers X dans L^2 . Par une démonstration analogue à la proposition 11, Les termes de la série sont majorés comme suit :

$$\begin{aligned} |\text{Cov}((X - X^h) \circ \theta_k, (X - X^h))| &\leq \text{Var}(X - X^h) \sqrt{\mathbb{P}(T^h > k)} \\ &\leq \text{Var}(X - X^h) \frac{\sqrt{\mathbb{E}(T^h)^{2+\varepsilon}}}{k^{1+\frac{\varepsilon}{2}}} \\ &<< \frac{\text{Var}(X - X^h)}{k^{1+\frac{\varepsilon}{2}}} \end{aligned}$$

Enfin, en introduisant cette inégalité dans l'équation 6.18 on obtient :

$$\tilde{\sigma}^2(X - X^h) << \text{Var}(X - X^h).$$

Ceci achève la démonstration. ■

En particulier, on peut dire que la vitesse de convergence mathématique, de cette méthode, pour le calcul des espérances des temps d'arrêts, est indépendante du pas de discrétisation.

6.5 Illustration informatique.

Nous avons voulu vérifier la validité du résultat obtenu ci-dessus par des simulations numériques. Nous avons repris l'exemple ci-avant, c'est à dire le système différentiel

$$\begin{cases} \frac{1}{2}u''(x) &= -1 \quad \forall x \in [-2, 2] \\ u(2) &= u(-2) = 0, \end{cases}$$

et on a calculé u au point 0. Rappelons que la solution de ce système est donnée par $u(x) = \mathbb{E}_x(\tau)$, avec τ est le temps de sortie du mouvement Brownien standard $w(t)$ de l'intervalle $[-2, 2]$. Les résultats obtenus montrent que la vitesse de convergence a été améliorée (voir figures 6.3 et 6.4).

- La représentation graphique ci-dessous représente la convergence du shift modifié et de Monte Carlo avec un pas de discrétisation $h = 0.1$.

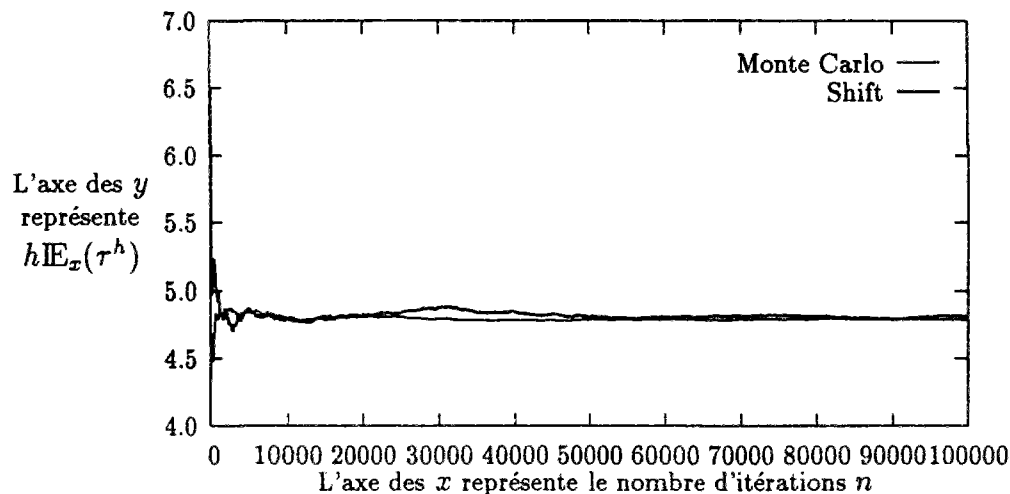


Figure 6.3

- La représentation graphique ci-dessous représente la convergence du shift modifié et de Monte Carlo avec un pas de discrétisation $h = 0.001$.

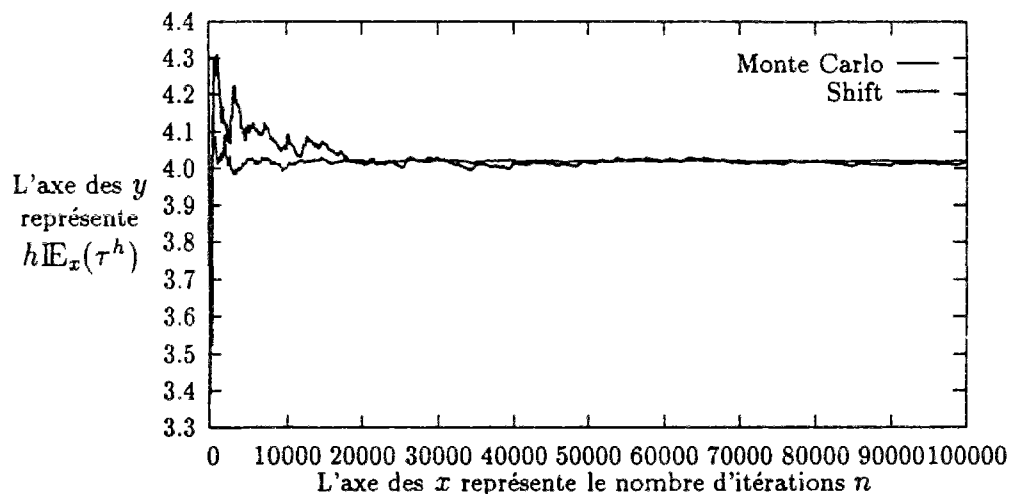


Figure 6.4

Nous voyons donc la méthode de décalage appliquée de la façon expliquée au paragraphe 6.4 et 6.5 apporte déjà une amélioration sensible vis-à-vis de la méthode de Monte Carlo "classique" pour le temps de calcul sur machine en raison de l'économie de tirage de gaussiennes à chaque itération.

Mais on peut espérer aller beaucoup plus loin comme nous allons le discuter maintenant. Contrairement à la méthode de Monte Carlo, la vitesse de convergence mathématique, pour le calcul de l'espérance d'une variable aléatoire X par la méthode du shift, dépend de la

simulation choisie. En effet, s'il existe deux algorithmes de simulation de X , représentés par deux fonctionnelles F et F' sur \mathbb{R}^N , il est clair que $\sigma^2(F)$ et $\sigma^2(F')$ peuvent avoir des ordres de grandeurs complètement différents. Ainsi, il faut penser à tester d'autres schémas de discrétisation et d'autres représentations du mouvement brownien. Par exemple, il est possible de simuler le mouvement brownien avec une procédure récursive.

6.6 Emploi de la méthode du shift sur d'autres représentations du mouvement brownien.

On peut montrer facilement à partir des propriétés suivantes :

- A t fixé, pour tous s , la variable aléatoire $w_{t+s} - w_s$ est une gaussienne centrée de variance t .
- Pour $0 \leq t_1 < t_2 < \dots < t_N$, les variables aléatoires $w_{t_1}, w_{t_2} - w_{t_1}, \dots, w_{t_N} - w_{t_{N-1}}$ sont indépendantes.

que pour $s < t$,

$$\mathcal{L}\left(w_{\frac{t+s}{2}}/w_s = x, w_t = y\right) = \mathcal{N}\left(\frac{x+y}{2}, \frac{t-s}{4}\right),$$

c'est à dire que, conditionnellement aux valeurs déjà prises par le brownien aux temps s et t , la loi du brownien au temps $\frac{t+s}{2}$ est une gaussienne de moyenne $\frac{x+y}{2}$ et de variance $\frac{t-s}{4}$. Ainsi, pour simuler une trajectoire brownienne sur l'intervalle $[0, 1]$ par exemple, on commence par tirer les points $w_0 = 0, w_1$ de loi $\mathcal{N}(0, 1)$, puis simuler $w_{\frac{1}{2}}$ par la propriété précédente et on fait alors de même pour obtenir $w_{\frac{1}{4}}, w_{\frac{3}{4}}, \dots$. On dispose ainsi d'une procédure récursive qui permet, à la profondeur m , d'obtenir les valeurs de $(w_{\frac{k}{2^m}})_{0 \leq k \leq 2^m}$. Nous renvoyons à O. Faure [15] pour une comparaison numérique sur la simulation du mouvement brownien par la méthode récursive et la méthode itérative.

Comment peut-on utiliser la méthode du shift ? Chaque réalisation fait intervenir des simulations de variables aléatoires gaussiennes centrées réduites. Le mouvement brownien admet donc une représentation sur l'espace $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mathcal{N}(0, 1)^{\otimes N})$, décrite par la procédure récursive. Supposons qu'on ait à simuler le mouvement brownien, sur l'intervalle $[0, 1]$ à une profondeur m , et qu'on ait donc utilisé les g_1, g_2, \dots, g_{2^m} premières coordonnées, alors la deuxième trajectoire "décalée" est calculée par la même procédure sur les coordonnées $(g_2, g_3, \dots, g_{2^m+1})$.

Toujours dans le même esprit : il est possible d'utiliser les développement de mouvement brownien sur la base de Haar, voir d'autres bases de $L^2[0, 1]$. En effet, soit $(\chi_n(t))_{n \geq 0}$ une base orthogonale de $L^2[0, 1]$, si l'on pose $\varphi_n(t) = \int_0^t \chi_n(s) ds$ alors pour toute suite $(g_n)_{n \geq 0}$ de variables aléatoires gaussiennes centrées réduites, considérée comme coordonnées de $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mathcal{N}(0, 1)^{\otimes N})$, la série $\sum_{n=0}^{\infty} g_n(\omega) \varphi_n(t)$ converge dans $C[0, 1]$ et dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$, $p \in [1, \infty[$, vers un mouvement brownien (voir [13]). Dans ce cas aussi le mouvement brownien admet une représentation sur $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N), \mathcal{N}(0, 1)^{\otimes N})$, décrite par la série ci-dessus. Comment peut-on utiliser la méthode du shift ? Supposons qu'on ait à

simuler le mouvement brownien, sur l'intervalle $[0, 1]$ par $\sum_{n=0}^N g_n(\omega)\varphi_n(t)$, avec une précision fixée au début, et qu'on ait donc utilisé les g_1, g_2, \dots, g_N premières coordonnées, alors la deuxième trajectoire "décalée" est donnée par $\sum_{n=0}^N g_{n+1}(\omega)\varphi_n(t)$. Ce sont, donc, les gaussiennes qui sont décalées et non la trajectoire. Par suite, on peut espérer que la corrélation entre les trajectoires soit moins importante.

La base de Haar est définie par : soit $\chi = 1_{[0, \frac{1}{2}[} - 1_{[\frac{1}{2}, 1[}$; si l'on pose $\chi_{m,k}(t) = 2^{\frac{m}{2}} \chi(2^m t - k)$, pour tout $t \in \mathbb{R}_+$, $m \in \mathbb{Z}$ et $k \in \mathbb{N}$, alors les fonctions $(\chi_{m,k})_{m \in \mathbb{Z}, k \in \mathbb{N}}$ forment une base de $L^2(\mathbb{R}_+)$. Nous renvoyons à N. Bouleau [13] pour plus de détails.

Des travaux sont en cours sur le sujet.

Partie II

Approximation des mesures invariantes

Chapitre 7

Calcul d'une intégrale par rapport à une mesure invariante.

Pour calculer l'intégrale d'une fonction f par rapport à une mesure μ , prise comme étant la mesure invariante d'une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$, on peut utiliser les théorèmes ergodiques pour les chaînes de Markov, à savoir sous des hypothèses convenables que

$$\int f d\mu = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(X_n). \quad (7.1)$$

Cette technique a été utilisée dans la résolution de problèmes physiques. Dans la première partie, on rappelle l'essentiel des résultats sur l'ergodicité des chaînes de Markov qui constituent le socle théorique de cette méthode. Au point de vue vitesse de convergence, nous rapporterons les observations retenues par les physiciens sur l'efficacité de cette méthode. La deuxième partie est consacrée à l'application de ce procédé au calcul d'exposants de Lyapunov d'une équation différentielle stochastique elliptique. Pour calculer cet exposant de Lyapunov λ , on commence par discrétiser l'équation puis on calcule directement l'exposant de l'équation obtenue (voir Pardoux et Pignol[40], et Talay[52]). Autrement dit on approche λ par $\lim_{n \rightarrow \infty} \frac{1}{n} \log \|\bar{X}_n^h\|$, avec $(\bar{X}_n^h)_{n \in \mathbb{N}}$ le processus approchant et h le pas de discrétisation. Par ailleurs, l'exposant de Lyapunov λ est aussi donné par une formule de type $\lambda = \int_{S^{d-1}} q(s) d\mu(s)$, où μ est la probabilité invariante d'un processus de Markov à valeurs dans la sphère unité S^{d-1} et q une fonction explicite. Ceci fournit un autre procédé de calcul. Une comparaison de la stabilité numérique de ces deux méthodes sera donnée à la fin de ce chapitre.

7.1 Présentation de la méthode ergodique.

Pour situer le cadre mathématique de la méthode ergodique, on rappellera quelques résultats sur la théorie des chaînes de Markov et on donnera, dans la dernière partie de ce paragraphe, des exemples où elle a été utilisée.

Notations et définitions : Désignons par $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov à valeurs dans un espace E , muni d'une tribu \mathcal{A} . On supposera que \mathcal{A} est séparable. On notera par $P(x, \cdot)$

la probabilité de transition de ce processus définie par :

$$P(x, A) = \mathbb{P}(X_{n+1} \in A / X_n = x)$$

Dans la littérature on montre sous des conditions diverses l'existence d'une mesure invariante μ vérifiant $\mu P = \mu$, c'est-à-dire

$$\forall A \in \mathcal{A}, \quad \mu(A) = \int_E P(x, A) \mu(dx).$$

μ est appelée la mesure invariante par P ou la mesure stationnaire du processus de Markov. Dans la suite nous donnons, brièvement, l'essentiel des résultats connus. Commençons par le cas où l'espace des états E est dénombrable.

7.1.1 Rappel sur la récurrence des chaînes de Markov à espace d'états dénombrable.

La probabilité de transition $P(x, \cdot)$ est, donc, définie par la donnée d'une matrice $P(x, y)$ vérifiant, $\sum_{y \in E} P(x, y) = 1 \forall x \in E$ et $P(x, y) > 0; \forall x, y \in E$. Cette matrice représente la probabilité de passage de l'état x à l'état y , $P(x, y) = \mathbb{P}(X_{n+1} = y / X_n = x)$. Si l'on note par $N_x = \sum_{n \in \mathbb{N}} 1_{\{X_n = x\}}$ le nombre de passage par x , on a alors le résultat suivant (cf. par exemple [46]).

Proposition 14 *Pour tout $x \in E$:*

- (a) *ou bien $\mathbb{P}(N_x < \infty / X_0 = x) = 1$; l'état x est alors dit transitoire.*
- (b) *ou bien $\mathbb{P}(N_x = \infty / X_0 = x) = 1$; l'état x est alors dit récurrent.*

■

La proposition précédente n'étudie les passages par un état y d'une chaîne de Markov partant de l'état x que lorsque $x = y$. Pour étudier le cas général on introduit l'hypothèse d'irréductibilité qui simplifie notablement la situation. Un sous-ensemble non vide A de l'espace des états E sera dit clos pour la matrice de transition P si, $\mathbb{P}(x, A) = 1 \forall x \in A$.

Définition 7 *Une matrice de transition P est dite irréductible si elle ne possède d'autre ensemble clos que E .*

■

A cet effet, il est nécessaire et suffisant que $\sum_{k=1}^{\infty} P^k(x, y) > 0$ pour tout $(x, y) \in E^2$, où $P^k(x, y) = \mathbb{P}(X_{n+k} = y / X_n = x)$. Introduisons la notion de chaînes récurrentes et de chaînes transitoires.

Proposition 15 *Pour toute matrice P irréductible, deux cas seulement sont possibles*

- (a) *soit $\mathbb{P}(N_y < \infty / X_0 = x) = 1$, quels que soient $x, y \in E$; la matrice P et la chaîne sont dites transitoires.*
- (b) *soit $\mathbb{P}(N_y = \infty / X_0 = x) = 1$, quels que soient $x, y \in E$; la matrice P et la chaîne sont dites récurrentes.*

■

Ces résultats montrent qu'une chaîne de Markov récurrente irréductible visite p.s. chacun des états de E une infinité de fois quel que soit son état initial. On montre qu'une telle matrice P admet une mesure invariante μ et que la fréquence du nombre de visites à y avant l'instant n , soit $\frac{1}{n} \sum_{m < n} 1_{\{X_m=y\}}$, est asymptotiquement proportionnel à $\mu(y)$.

Proposition 16 *Toute matrice récurrente irréductible P admet une mesure positive invariante μ sur E ($\mu P = P$), et une seule à une constante multiplicative près.* ■

L'intérêt de la mesure invariante μ introduite par la proposition précédente tient notamment au résultat suivant qui découle du théorème ergodique.

Proposition 17 *Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov récurrente irréductible et soit P sa matrice de transition. Si P admet une probabilité invariante μ sur E , alors*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} f(X_n) = \int_E f(x) \mu(dx) \quad \mathbb{P}(d\omega) - p.s.$$

quelle que soit la fonction $f \in L^1(\mu)$ ou f positive. En général si μ désigne "la" mesure positive P -invariante sur E alors

$$\lim_{N \rightarrow \infty} \frac{\sum_{n=0}^{N-1} f(X_n)}{\sum_{n=0}^{N-1} g(X_n)} = \frac{\int_E f(x) \mu(dx)}{\int_E g(x) \mu(dx)} \quad \mathbb{P}(d\omega) - p.s.$$

pour toute $f \in L^1(\mu)$ et g positive. ■

7.1.2 Rappel sur la récurrence des chaînes de Markov à espace d'états non dénombrable.

Dans le cas général où E n'est pas dénombrable nous donnerons des hypothèses de récurrence, souvent utilisée dans des textes élémentaires (voir par exemple [46], [38], [24] et [36]), sous lesquelles certaines chaînes dites de Harris vérifient les résultats de la proposition précédente. Soit ν une mesure de probabilité sur E . Donnons la définition d'une chaîne ν -irréductible.

Définition 8 *Une chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ associée à une probabilité de transition P est dite ν -irréductible si, pour tout $A \in \mathcal{A}$ tel que $\nu(A) > 0$, on a $\sum_{k=1}^{\infty} P^k(x, A) > 0$ pour tout $x \in E$, où $P^k(x, A) = \mathbb{P}(X_k \in A / X_0 = x)$.* ■

Introduisons la notion de chaînes récurrentes et de chaînes transitoires.

Proposition 18 *Pour toute probabilité de transition P , ν -irréductible, deux cas seulement sont possibles*

- (a) *soit E est le réunion croissante d'ensembles transitoires, c'est-à-dire il existe une suite croissante A_n telle que $\cup_{n \in \mathbb{N}} A_n = E$ et $\mathbb{P}(N_{A_n} < \infty / X_0 = x) = 1$ pour tous $x \in E$ et $n \in \mathbb{N}$; la probabilité de transition P et la chaîne sont dites transitoires.*

- (b) soit il existe une mesure positive σ -finie μ , P -invariante telle que $\nu \ll \mu$ vérifiant : $\forall A \in \mathcal{A}$ tel que $\mu(A) > 0$ on a, $\mathbb{P}(N_A = \infty / X_0 = x) = 1 \forall x \in E$; la chaîne est dite *récurrente au sens de Harris ou de Harris*.

■

Pour une chaîne de Harris, il existe une famille finie d'ensembles $(C_i)_{1 \leq i \leq d}$ deux à deux disjoints telle que :

- (i) Le complémentaire de la réunion des $(C_i)_{1 \leq i \leq d}$ est de mesure nulle.
- (ii) $P(x, C_1) = 1 \forall x \in C_d$ et $P(x, C_{j+1}) = 1 \forall x \in C_j$ et $\forall 1 \leq j \leq d-1$.

d est appelé la période de $(X_n)_{n \in \mathbb{N}}$. Quand $d = 1$ la chaîne est dite apériodique.

Théorème 22 (Orey) Soient $(X_n)_{n \in \mathbb{N}}$ une chaîne de Harris apériodique, alors pour toutes ν_1 et ν_2 mesures de probabilités sur \mathcal{A}

$$\lim_{n \rightarrow \infty} \|(\nu_1 - \nu_2)P^n\| = 0.$$

En particulier s'il existe une probabilité invariante alors elle est unique et, pour toute mesure de probabilité ν , sur \mathcal{A}

$$\lim_{n \rightarrow \infty} \|\nu P^n - \mu\| = 0.$$

La norme $\|\nu\|$ désigne la norme totale de la mesure signée ν .

■

Dans certain cas, la vitesse de convergence de ces limites est géométrique (Doebelin récurrente). Il est intéressant de remarquer que pour un espace d'état E dénombrable, les chaînes récurrentes irréductibles sont toutes des chaînes de Harris.

7.1.3 Utilisation de la méthode par les physiciens.

Pour calculer des grandeurs relatives à l'équilibre thermodynamique en physique statistique (voir [3],[10],[25]) ou pour calculer des grandeurs attachées aux modèles de théorie de jauge sur réseaux en mécanique quantique (voir [3],[25],[30],[45]) on est fréquemment dans la situation suivante :

- On veut calculer $\int_E f(x) d\mu(x)$ où E est un espace fini discret.
- Pour chaque $x \in E$ les nombres $\mu(x)$, $f(x)$ sont accessibles au calcul.
- Mais le cardinal de E est vertigineux, c'est l'espace des configurations, et on ne dispose pas d'une structure de données permettant de décrire simplement les ensembles $\{x \in E : \mu(x) > \varepsilon\}$.

Dans une telle situation les physiciens utilisent beaucoup les méthodes de Monte Carlo selon le principe suivant (voir [25]) : On choisit une probabilité de transition $P(x, y)$ de E dans E admettant μ comme probabilité invariante. Partant d'une loi μ_0 sur E on construit la chaîne de Markov X_n de loi initiale μ_0 et de matrice de transition P et on calcule :

$$\int_E f(x) d\mu(x) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(X_n) \quad (7.2)$$

en supposant que les hypothèses de la proposition 17 des chaînes de Markov s'appliquent. On comprend donc que l'algorithme définie par l'équation (7.2) réalise en fait deux choses simultanément. Il approche μ par $\mu_0 P^n$ et il calcule $\int_E f(x) d\mu(x)$ par une moyenne de Cesaro le long d'une trajectoire de la chaîne de Markov stationnaire. La mesure initiale est généralement choisie égale à la masse de Dirac en un point x_0 . Il convient de souligner qu'il y a toujours plusieurs façons de construire la probabilité de transition P . D'autre part, il n'est pas nécessaire pour définir P de passer en revue tous les points x de E mais seulement de dire comment (par quels appels à la fonction random) on passe d'un x connu à un certain y .

La faiblesse de ce procédé est la vitesse de convergence. En effet, si l'on parcourt la littérature, notamment sur la chromodynamique quantique sur réseaux on s'aperçoit que les physiciens ont rencontré des vitesses de convergence de l'algorithme (7.2) extrêmement variables (voir [25]) suivant les situations. Ils obtiennent souvent des résultats significatifs avec une précision de 10^{-2} avec 500 itérations seulement. Parfois même il y a stabilisation au bout de 30 itérations. Au contraire pour certains problèmes ils ressentent la nécessité de procéder à 21000 itérations représentant 17 heures sur le CDC 76000 du CERN.

A quoi sont dues de telles vitesses de convergence, surprenantes en simulation aléatoire ? La vitesse de convergence est contrôlée par la variance qui est égale $\int (f - \int f d\mu)^2 d\mu$. Dans certaines applications en physique ce paramètre est petit parce que f est presque constante sur les points où μ a une valeur significative.

Nous présentons dans le paragraphe suivant un exemple d'une situation analogue.

7.2 Exposant de Lyapunov.

Nous considérons l'équation bilinéaire stochastique sur \mathbb{R}^d , donnée au sens de Stratonovich par :

$$dx(t) = Ax(t)dt + \sum_{i=1}^m B_i x(t) \circ dw_t^i \quad (7.3)$$

où $\{w_1(t), \dots, w_m(t); t \geq 0\}$ est une famille de processus de Wiener standard, mutuellement indépendants, définis sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_t)_{t \geq 0}$. $x(t) \in \mathbb{R}^d$ et A, B_1, \dots, B_m sont des $d \times d$ matrices. Pour une variable aléatoire x_0 , \mathcal{F}_0 -mesurable, on notera par $x(\cdot, x_0)$ la solution de l'équation (7.3) vérifiant la condition initiale $\mathbb{P}\{x(\cdot, x_0) = x_0\} = 1$.

Dans une large mesure, le comportement asymptotique de la solution de l'équation (7.3) est caractérisé par son exposant de Lyapunov. En effet, si la limite ci-dessous existe, elle est indépendante du hasard et de la condition initiale.

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \log |x(t, x_0)| \quad \mathbb{P}(d\omega) - p.s. \quad (7.4)$$

où $|\cdot|$ désigne la norme euclidienne sur \mathbb{R}^d . Le réel λ est appelé exposant de Lyapunov de l'équation (7.3). Si $\lambda < 0$ le système est dit stable, auquel cas la solution de l'équation (7.3) décroît vers 0 à une vitesse exponentielle.

Il s'ensuit que la projection de la solution $x(t, x_0)$ sur la sphère unité $S^{d-1} = \{x \in \mathbb{R}^d \text{ tel que } |x| = 1\}$, notée par :

$$s_0 = \frac{x_0}{|x_0|}, \quad s(t, s_0) = \frac{x(t, x_0)}{|x(t, x_0)|},$$

est un processus de diffusion sur S^{d-1} (voir N. Ikeda et S. Watanabe [28], page 109), qui vérifie l'équation

$$\begin{cases} s(0) &= s_0 \\ ds(t) &= h_0(s)dt + \sum_{i=1}^m h_i(s) \circ dw_t^i. \end{cases} \quad (7.5)$$

Le champ de vecteur $h_0(s), h_1(s), \dots, h_m(s)$ est donné par

$$\begin{aligned} h_0(s) &= h(A, s) \\ h_i(s) &= h(B_i, s) \quad 1 \leq i \leq m, \end{aligned}$$

où $\forall s \in S^{d-1}$ et $\forall C$ une $d \times d$ matrice

$$h(C, s) = (C - (s'Cs)I)s. \quad (7.6)$$

Dans l'équation ci-dessus (7.6), on a noté s' le transposé de s et par I la matrice identité.

Dans la suite du chapitre, on supposera que le champ de vecteurs $\{h_0, \dots, h_m\}$, défini dans le système (7.5), vérifie l'hypothèse suivante :

$$(H) \quad \dim L.A.\{h_0, \dots, h_m\}(s) = d - 1 \quad \forall s \in S^{d-1}, \quad (7.7)$$

où $L.A.\{h_0, \dots, h_m\}$ est l'algèbre de Lie engendré par le champ de vecteurs $\{h_0, \dots, h_m\}$. Sous cette hypothèse, Arnold, Oeljeklaus et Pardoux [5] ont établi le résultat suivant.

Théorème 23 *Sous l'hypothèse (H) et si $x_0 \neq 0$ on a :*

(a) *L'exposant de Lyapunov λ existe, à savoir que*

$$\forall x_0 \in \mathbb{R}^d - \{0\} \quad \lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \log |x(t, x_0)| \quad p.s.$$

(b) *En plus le processus $(s(t))_{t \geq 0}$ admet une unique probabilité invariante μ et l'exposant de Lyapunov λ est donné par*

$$\lambda = \int_{S^{d-1}} Q(s) d\mu(s), \quad (7.8)$$

où

$$Q(s) := s'As + \frac{1}{2} \sum_{i=1}^m \{s'A_i^2s + |A_i s|^2 - 2(s'A_i s)^2\}. \quad (7.9)$$

■



En parcourant la littérature, on ne trouve pas de formules explicites qui donnent la valeur exacte de λ . Sauf dans certains cas, pour des systèmes dynamiques en dimension 2, on trouve des résultats dans cette direction, en particulier pour l'oscillateur harmonique aléatoire (voir Arnold, Papanicolaou et Wihstutz [6], Pinsky [43], Pardoux et Wihstutz [41], Pinsky et Wihstutz [44] et Kleimann [31]). En outre, dans la plus part de ces travaux, on a un système dynamique qui dépend d'un paramètre ε et on étudie le comportement asymptotique de l'exposant de Lyapunov lorsque ε tend 0. Il est donc intéressant d'approcher numériquement l'exposant de Lyapunov λ .

Dans le paragraphe suivant, nous rappelons les résultats obtenus par D. Talay sur l'approximation de l'exposant de Lyapunov d'un système différentiel stochastique bilinéaire; il approche λ par $\lim_{n \rightarrow \infty} \frac{1}{n} \log \|\bar{x}_n^h\|$ où $(\bar{x}_n^h)_{n \in \mathbb{N}}$ le processus approchant et h le pas de discrétisation. Pour notre part, nous proposons une autre méthode de simulation basée sur l'approximation de la mesure invariante μ . En effet, l'exposant de Lyapunov λ peut être approcher par $\int_{S^{d-1}} Q(s) d\bar{\mu}^h(s)$ où $\bar{\mu}^h$ est la mesure invariante du processus approximant et h le pas de discrétisation. Une comparaison de la stabilité numérique de ces deux méthodes est donnée à la fin de ce paragraphe.

7.2.1 Approximation des exposants de Lyapunov.

A ce sujet, D. Talay [52] a proposé un algorithme de calcul de stabilité des systèmes différentiels stochastiques bilinéaires. Il a également étudié le cas non linéaire avec Axel Grorud (voir [20]). Il a montré que, pour plusieurs schémas de discrétisation $(\bar{x}_p^h, p \in \mathbb{N})$, les exposants de Lyapunov associés

$$\bar{\lambda}^h := \lim_{p \rightarrow \infty} \frac{1}{ph} \log |\bar{x}_p^h| \quad p.s. \quad (7.10)$$

sont bien définis, et il a estimé en fonction du pas de discrétisation h l'erreur théorique

$$\left| \lambda - \lim_{p \rightarrow \infty} \frac{1}{ph} \log |\bar{x}_p^h(x_0)| \right|$$

où x_0 est une condition initiale non nulle arbitraire. Plus précisément, en supposant que le générateur infinitésimal de l'équation (7.3) est strictement elliptique, il a montré que les schémas d'Euler et de Milstein vérifient (voir [52]) :

$$|\lambda - \bar{\lambda}^h| = O(h)$$

et que les schémas du second ordre pour l'approximation de $\mathbb{E}f(x(t))$ en temps fini vérifient (voir [52]) :

$$|\lambda - \bar{\lambda}^h| = O(h^2)$$

Dans la suite on se restreindra au schéma d'Euler "classique", et on remplacera donc les accroissements du mouvement brownien par des gaussiennes indépendantes. Le système dynamique (7.3) étant équivalent au système :

$$dx(t) = \left(A + \frac{1}{2} \sum_{i=1}^m B_i^2 \right) x(t) dt + \sum_{i=1}^m B_i x(t) dw_t^i,$$

il résulte que la discrétisation de l'équation différentielle stochastique (7.3) par la méthode d'Euler est donnée par :

$$\begin{cases} \bar{x}_0^h = x_0 \\ \bar{x}_{p+1}^h = \bar{x}_p^h + \tilde{A}\bar{x}_p^h + \sum_{i=1}^m B_i \bar{x}_p^h g_{p+1}^i \sqrt{h} \end{cases} \quad (7.11)$$

où $\tilde{A} := A + \frac{1}{2} \sum_{i=1}^m B_i^2$ et $(g_p^j)_{p,j \in \mathbb{N}}$ est une famille de variables aléatoires gaussiennes centrées réduites. Précisons les hypothèses des résultats obtenus par D. Talay dans le cas du schéma d'Euler.

Proposition 19 *Sous l'hypothèse (H) et si $x_0 \neq 0$ on a :*

(a) *Pour h assez petit, il existe $\bar{\lambda}^h \in \mathbb{R}$:*

$$\forall x_0 \in \mathbb{R}^d - \{0\} \quad \bar{\lambda}^h = \lim_{p \rightarrow \infty} \frac{1}{ph} \log |\bar{x}_p^h| \quad p.s.$$

(b) *et même :*

$$\forall x_0 \in \mathbb{R}^d - \{0\} \quad \bar{\lambda}^h = \lim_{p \rightarrow \infty} \frac{1}{ph} \mathbb{E} \log |\bar{x}_p^h|.$$

(c) *En plus, le processus $\bar{s}_p^h := \frac{\bar{x}_p^h}{|\bar{x}_p^h|}$ est ergodique, il admet, donc, une unique probabilité invariante $\bar{\mu}^h$ et l'exposant de Lyapunov $\bar{\lambda}^h$ du système (7.11) est donné par*

$$\bar{\lambda}^h = \int_{S^{d-1}} \bar{Q}^h(s) d\bar{\mu}^h(s), \quad (7.12)$$

où

$$\bar{Q}^h(s) := \frac{1}{h} \mathbb{E} \log |\bar{x}_1^h(s)| \quad (7.13)$$

Preuve : La démonstration est basé sur un théorème de Furstenberg. Voir proposition 4.1 dans l'article de D. Talay[52]. ■

Dans la référence [52], on donne une estimation de la vitesse de convergence.

Théorème 24 *Si l'on suppose que le générateur infinitésimal du processus $(s(t))_{t \geq 0}$, défini par l'équation (7.5), est strictement elliptique, autrement dit il existe $\alpha > 0$ tel que $\forall x \in S^{d-1}$ et $\forall \xi \in T_{S^{d-1}}(x)$*

$$\sum_{i=1}^m (h(B_i, x), \xi)^2 \geq \alpha |\xi|^2,$$

alors on a $|\lambda - \bar{\lambda}^h| = O(h)$. ■

Pour démontrer le théorème ci-dessus, on utilise le résultat suivant (voir [52]).

Lemme 11 *Si l'on suppose que le générateur infinitésimal du processus $(s(t))_{t \geq 0}$, défini par l'équation (7.5) est strictement elliptique alors, $\forall x_0 \in \mathbb{R}^d - 0$ et pour toute fonction f de classe C^∞ , $f : S^{d-1} \rightarrow \mathbb{R}$,*

$$\left| \int_{S^{d-1}} f(s) d\mu(s) - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{p=1}^N f(\bar{s}_p^h(\frac{x_0}{|x_0|})) \right| = O(h) \quad p.s.$$

où $x_0 \in \mathbb{R}^d$ est une condition initiale non nulle arbitraire. ■

En effet, en supposant que le générateur infinitésimal est strictement elliptique, on déduit que l'hypothèse (H) est vérifiée et on obtient, donc, les résultats (a), (b) et (c) de la proposition 19. Ceci nous permet de décomposer l'erreur théorique $\lambda - \bar{\lambda}^h$ comme suit :

$$\lambda - \bar{\lambda}^h = \int_{S^{d-1}} Q(s)(d\mu(s) - d\bar{\mu}^h(s)) + \int_{S^{d-1}} (Q(s) - \bar{Q}^h(s))d\bar{\mu}^h(s).$$

Par suite la preuve consiste essentiellement à vérifier que les opérateurs de transition des processus $(\bar{s}_p^h)_{p \in \mathbb{N}}$ approchent celui de la chaîne $(s(ph))_{p \in \mathbb{N}}$ et d'estimer l'écart entre les lois invariantes de $(\bar{s}_p^h)_{p \in \mathbb{N}}$ et de $(s(t))_{t \geq 0}$.

Les résultats énoncés ci-avant, nous suggèrent deux algorithmes d'approximation de l'exposant de Lyapunov d'un système dynamique différentiel. En effet, soit on approche λ par $\bar{\lambda}^h$ avec :

$$\bar{\lambda}^h = \int_{S^{d-1}} \bar{Q}^h(s) d\bar{\mu}^h(s) = \lim_{p \rightarrow \infty} \frac{1}{ph} \log |\bar{x}_p^h| \quad p.s. \quad (7.14)$$

auquel cas l'erreur théorique $|\lambda - \bar{\lambda}^h| = O(h)$, soit on l'approche par $\tilde{\lambda}^h$ où :

$$\tilde{\lambda}^h = \int_{S^{d-1}} Q(s) d\bar{\mu}^h(s) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{p=1}^N Q(\bar{s}_p^h) \quad p.s. \quad (7.15)$$

auquel cas l'erreur théorique est donnée par le lemme 11, $|\lambda - \tilde{\lambda}^h| = O(h)$. Il est donc intéressant de comparer la stabilité numérique des deux algorithmes sur des équations différentielles stochastiques dont on connaît explicitement l'exposant de Lyapunov.

7.2.2 Résultats numériques

L'utilisation de ces deux algorithmes pose un problème pratique important, qui reste ouvert. En effet, pour calculer $\bar{\lambda}^h$ (respectivement $\tilde{\lambda}^h$) on a besoin de déterminer le nombre d'itérations p_{max} (respectivement N_{max}) à partir duquel la quantité $\frac{1}{p_{max}h} \log |\bar{x}_{p_{max}}^h|$ (respectivement $\frac{1}{N_{max}} \sum_{p=1}^{N_{max}} Q(\bar{s}_p^h)$) est assez proche de la valeur limite $\bar{\lambda}^h$ (respectivement $\tilde{\lambda}^h$). Il est donc important d'estimer la vitesse de convergence des deux suites, $\frac{1}{ph} \log |\bar{x}_p^h|$ et $\frac{1}{N} \sum_{p=1}^N Q(\bar{s}_p^h)$, en fonction du nombre d'itérations. Dans la référence [53], on explique qu'une détermination de la vitesse de convergence nécessite une résolution d'une équation aux dérivées partielles dont la complexité est comparable à celle de l'équation de Fokker-Planck stationnaire et que cette équation est paramétrée par le terme que nous voulons calculer.

Plus précisément, si $(x_t)_{t \geq 0}$ est un processus ergodique de loi invariante μ et si l'on pose $\tilde{f} = f - \int f d\mu$ on a alors, sous certaines conditions : $\lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \int_0^T \tilde{f}(x_s) ds = \mathcal{N}(0, V(f))$ où $V(f) = 2 \int \tilde{f}(x) v(x) d\mu(x)$ et $v(x)$ est la solution de l'équation de Poisson $Lv = -\tilde{f}$ (voir Bhattacharya[8]).

Ainsi, nous n'avons pas actuellement de méthode plus fiable que d'attendre une stabilisation numérique de la suite et d'assimiler la valeur stable à sa limite. Par ailleurs, il est important de remarquer que le calcul direct de la formule $\frac{1}{ph} \log |\bar{x}_p^h|$ pose des problèmes numériques graves. En effet, soit $\bar{\lambda}^h$ est positif auquel cas la solution \bar{x}_p^h diverge vers l'infini à une vitesse exponentielle, soit $\bar{\lambda}^h$ est négatif auquel cas elle converge vers 0 et à une vitesse exponentielle aussi. Cependant, si l'on note par :

$$\begin{cases} \bar{y}_0^h &:= \bar{s}_0^h \\ \bar{y}_{p+1}^h &:= \bar{s}_p^h + \tilde{A} \bar{s}_p^h + \sum_{i=1}^m B_i \bar{s}_p^h g_{p+1}^i \sqrt{h} \quad \forall p \geq 1 \end{cases} \quad (7.16)$$

on a alors $\frac{\bar{y}_p^h}{\|\bar{y}_p^h\|} = s_p^h$ et $\|\bar{x}_p^h\| = \prod_{k=1}^p \|\bar{y}_k^h\|$, $\forall p \in \mathbb{N}$ (pour démontrer ces résultats, il suffit de remplacer dans l'équation (7.16) \bar{s}_p^h par $\frac{\bar{x}_p^h}{\|\bar{x}_p^h\|}$). Il est donc préférable d'utiliser l'algorithme donné par la formule :

$$\bar{\lambda}^h = \lim_{N \rightarrow \infty} \frac{1}{Nh} \sum_{p=1}^N \log \|\bar{y}_p^h\|. \quad (7.17)$$

Maintenant, nous allons comparer les deux algorithmes (7.15) et (7.17) sur des exemples dont on sait calculer les exposants de Lyapunov.

Exemple : Considérons le système dynamique dans \mathbb{R}^2 donné par :

$$dx(t) = Ax(t) + \sigma Bx(t) \circ dw(t),$$

où $(w(t))_{t \geq 0}$ est un processus de Wiener standard sur \mathbb{R} , $\sigma \in \mathbb{R}$, A est une matrice carrée de la forme $A := \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, avec a et $b \in \mathbb{R}$, et $B := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. D'après [52], l'exposant de Lyapunov de ce système est donné par la formule ci-dessous,

$$\lambda = \frac{1}{2}(a+b) + \frac{1}{2}(a-b) \frac{\int_0^{2\pi} \cos(2\theta) \exp(\frac{a-b}{2\sigma^2} \cos(2\theta)) d\theta}{\int_0^{2\pi} \exp(\frac{a-b}{2\sigma^2} \cos(2\theta)) d\theta}.$$

Si l'on prend $a := 1$, $b := -2$ et $\sigma := 10$ on trouve $\lambda := -0.489 < 0$. Le système est donc stable.

Pour l'algorithme (7.17) et avec des pas de discrétisation de plus en plus fins on a obtenu les résultats suivants :

- Prenons $h := 0.1$. La représentation graphique ci-dessous illustre la convergence de la suite, $\frac{1}{10^{-1}N} \sum_{p=1}^N \log \|\bar{y}_p^{0.1}\|$, vers sa valeur limite $\bar{\lambda}^{0.1}$.

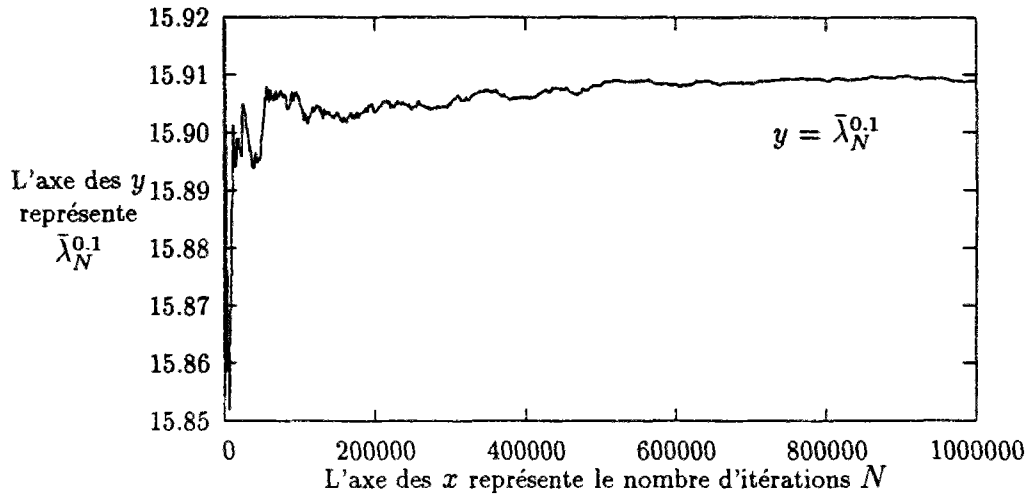


Figure 7.1

- Prenons $h := 0.001$. La représentation graphique ci-dessous illustre la convergence de la suite, $\frac{1}{10^{-3}N} \sum_{p=1}^N \log ||\tilde{y}_p^{0.001}||$, vers sa valeur limite $\bar{\lambda}^{0.001}$.

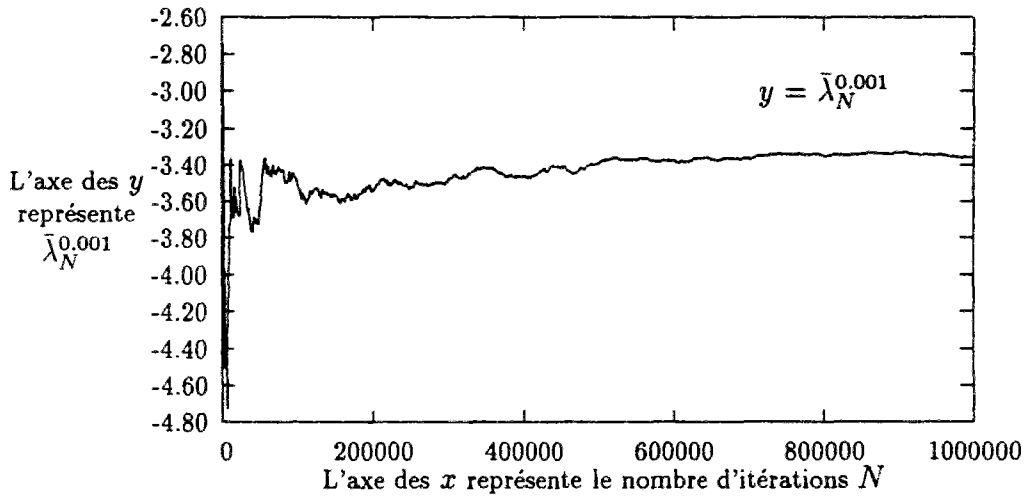


Figure 7.2

On remarque que pour un pas discrétisation $h = 0.1$, l'algorithme (7.17) laisserait penser que le système est fortement instable. Pour $h = 0.001$, il laisserait penser que le système est beaucoup plus stable qu'il ne l'est. L'utilisation de cet algorithme demande beaucoup de prudence. En effet, à chaque pas de discrétisation h correspond une chaîne de Markov dont l'exposant de Lyapunov $\bar{\lambda}^h$ peut être très différent de λ , il suffit que la constante sous-jacente de "O" de l'erreur soit grande. Ceci nous a conduit à faire une étude numérique plus fine sur la vitesse de convergence de $\bar{\lambda}^h$ vers λ en fonction du pas de discrétisation h . Pour cet

exemple on a été amené à prendre $h = 10^{-4}$ et voir même 10^{-5} pour obtenir des résultats acceptables. Par ailleurs, il est intéressant de noter que le nombre d'itérations N dépend énormément du pas de discrétisation choisi. Enfin, pour un temps d'intégration final $Nmax$, vérifiant $Nmaxh = 100$, les résultats obtenus se résument dans le tableau ci-dessous (pour h donné, $Nmax$ est le nombre d'itérations N dans la suite $\frac{1}{Nh} \sum_{p=1}^N \log \|\bar{y}_p^h\|$ pour lequel on a arrêté le calcul).

h	0.1	0.01	0.001	0.000 1	0.000 01
$\bar{\lambda}^h$	15.899	-11.865	-3.514	-0.895	-0.544
$ \lambda - \bar{\lambda}^h $	16.388	11.376	3.025	0.406	0.055
$Nmax$	1 000	10 000	100 000	1 000 000	10 000 000

Tableau 7.1

Ceci confirme les explications données ci-dessus sur les phénomènes de convergence trompeurs. Dans [52], l'auteur a étudié l'approximation d'un exposant de Lyapunov d'une équation différentielle stochastique bilinéaires via plusieurs schéma de discrétisation. Pour le même exemple, il a observé le même phénomène en utilisant un schéma de Milshtein. Les résultats théoriques et numériques étaient plus satisfaisant, lorsqu'il a choisi un schéma de discrétisation de second ordre (voir D. Talay [52]). En effet, il montre que pour un tel schéma l'erreur théorique $|\lambda - \bar{\lambda}^h| = O(h^2)$ et il obtient pour le même exemple, en prenant $Nh = 100$, $\bar{\lambda}^h = -0.35$ avec $h = 10^{-3}$ et $\bar{\lambda}^h = -0.50$ avec $h = 10^{-4}$.

Considérons, maintenant, l'algorithme (7.15) à savoir que : $\tilde{\lambda}^h = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{p=1}^N Q(\bar{s}_p^h)$. Sur le même exemple et avec des pas de discrétisation de plus en plus fin on a obtenu les résultats suivants.

- Prenons $h := 0.1$. La représentation graphique ci-après illustre la convergence de la suite, $\frac{1}{N} \sum_{p=1}^N Q(\bar{s}_p^{0.1})$, vers sa valeur limite $\tilde{\lambda}^{0.1}$.

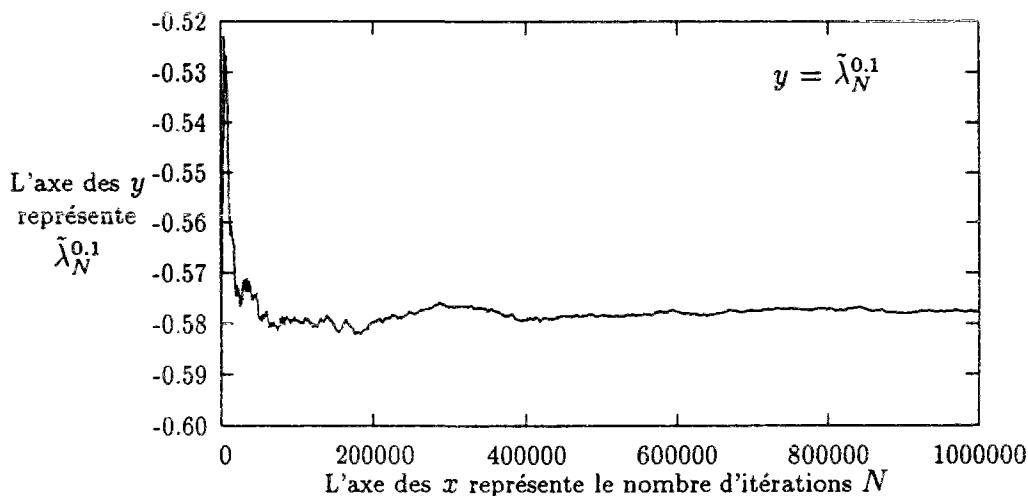


Figure 7.3

- Prenons $h := 0.001$. La représentation graphique ci-dessous illustre la convergence de la suite, $\frac{1}{N} \sum_{p=1}^N Q(\bar{s}_p^{0.001})$, vers sa valeur limite $\bar{\lambda}^{0.001}$.

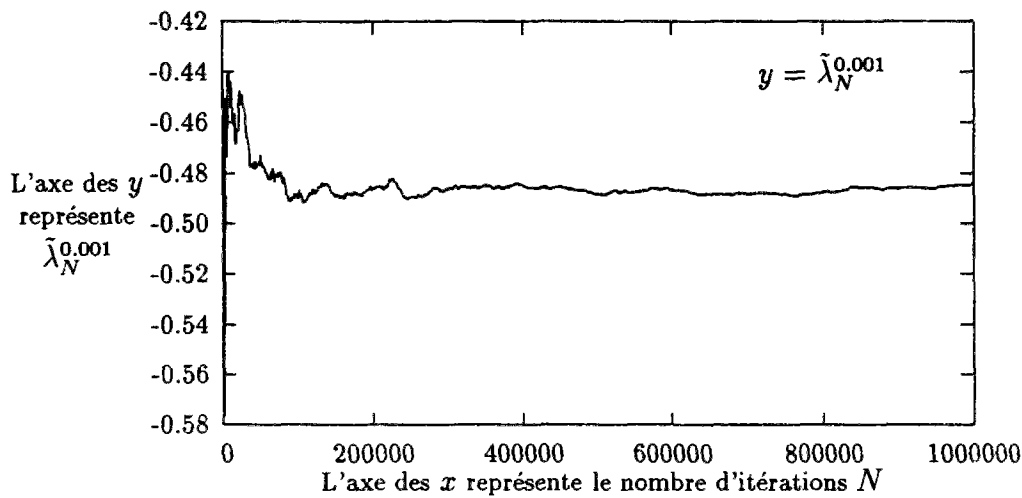


Figure 7.4

De même que pour l'algorithme (7.17), voir Tableau 7.1, on a fait des statistiques sur la vitesse de convergence en fonction du pas de discrétisation h : si l'on prend comme critère sur le nombre d'itérations, $Nh = 100$, on obtient les résultats suivants.

h	0.1	0.05	0.01	0.005	0.001
$\bar{\lambda}^h$	-0.593	-0.575	-0.469	-0.483	-0.4885
$ \lambda - \bar{\lambda}^h $	0.104	0.08	0.019	0.005	0.0004
N	1 000	2 000	10 000	20 000	100 000

Tableau 7.2

Il est clair que d'un point de vue numérique, ces résultats sont plus satisfaisants. On constate que pour des pas de discrétisation $h \leq 0.001$, on a des erreurs absolues de l'ordre de 10^{-2} voir même 10^{-3} .

Cependant le problème de détermination de la vitesse de convergence d'un algorithme ergodique reste ouvert. Il est donc intéressant de trouver une estimation a priori (théorique ou numérique) des variances des erreurs normalisées de type

$$\frac{1}{\sqrt{t}} \int_0^t \left(f(X_s) - \int f d\mu \right) ds$$

où $(X_t)_{t \geq 0}$ est un processus ergodique de loi invariante μ .

Enfin une autre direction de recherche, pour le calcul d'une intégrale par rapport à une mesure obtenue comme mesure limite invariante, pourrait être d'approcher le processus $(X_t)_{t \geq 0}$ par un processus de Markov à sauts obtenu par discrétisation de l'espace d'états.

7.3 Conclusion.

Sous certaines hypothèses un processus de Markov $(X_t)_{t \geq 0}$ admet une unique mesure invariante μ . Par ailleurs, on a une fonction f , par exemple mesurable bornée, et on cherche à calculer $\int f d\mu$. Une façon d'approximer la loi limite est de la remplacer par la mesure invariante $\bar{\mu}^h$ d'un processus approximant. Ceci revient à considérer $\int f d\bar{\mu}^h$. Pour l'erreur théorique $|\int f d\mu - \int f d\bar{\mu}^h|$ et dans le cas où $(X_t)_{t \geq 0}$ est solution d'une équation différentielle stochastique sur \mathbb{R}^d , nous renvoyons à D. Talay (voir [51], [54], [53]), où l'on trouve une estimation de cette erreur pour plusieurs schémas de discrétisation. En général, et lorsque la chaîne approximante $(\bar{X}_k^h)_{k \in \mathbb{N}}$ est ergodique, on approche $\int f d\bar{\mu}^h$ par

$$\frac{1}{N} \sum_{k=1}^N f(\bar{X}_k^h),$$

pour N assez grand. Le choix de N et de h est délicat ; on ne connaît pas de résultat général permettant de faire un bon choix pour une précision donnée. Mais il convient de remarquer que les résultats numériques obtenus dans le cas des exposants de Lyapunov sont satisfaisant.

Bibliographie

- [1] M. B. Alaya. Sur la méthode du shift en simulation. N. Bouleau et D. Talay, eds, *Probabilités numériques*, volume 10, chapitre 2, pages 61–66. INRIA, 1992.
- [2] M. B. Alaya. On the simulation of expectations of random variables depending on a stopping time. *Stochastic Analysis and Applications*, 1993. À paraître.
- [3] R. Alcouffe, ed. *Monte Carlo Methods and applications in Neutronics, Phonotics and Statistical Physics*, n° 240 , Lecture Notes in Physics. Springer-Verlag, 1985.
- [4] L. Stuart Anderson. Random number generators on vector supercomputers and other advanced architectures. *SIAM Rev*, 32(2) :221–251, 1990.
- [5] L. Arnold, E. Oeljeklaus, et E. Pardoux. Almost sure and moment stability for linear Ito equations. L. Arnod et V. Wihstutz, eds, *Lyapunov Exponents*, n° 1186 , Lecture Notes in Mathematics. Springer, 1986.
- [6] L. Arnold, G. Papanicolaou, et V. Wihstutz. Asymptotic analysis of the Lyapunov exponent and rotation number of the random oscillator and applications. *SIAM J. Appl. Math.*, 46(3) :427–450, 1986.
- [7] E. Berger. An almost sure invariance principle for stationary sequences of Banach space valued random variables. *Probab. th. rel. Fields*, 84 :161–201, 1990.
- [8] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 60 :185–201, 1982.
- [9] P. Billingsley. *Convergence of Probability Measures*. Wiley, New-York, 1968.
- [10] K. Binder, ed. *Monte Carlo Methods in Statistical Physics*, volume 7, *Topic Cur. Phys.* Springer, 1986.
- [11] N. Bouleau. *Probabilités de l'ingénieur, variables aléatoires et simulation*. Hermann, Paris, 1986.
- [12] N. Bouleau. On effective computation of expectations in large or infinite dimension. *Journal of Computational and Applied Mathematics*, 31 :23–34, 1990. North-Holland.
- [13] N. Bouleau. On numerical integration by the shift and application to Wiener space. *Acta Applicandae Mathematicae*, 25(3) :201–220, 1991.

- [14] N. Dunford et J. T. Schwartz. *Linear Operators*. part I. Interscience, New York, 1967.
- [15] O. Faure. *Simulation du mouvement brownien et des diffusions*. Thèse de l'école nationale des ponts et chaussées, 1992.
- [16] W. Feller. *An introduction to Probability Theory and its Applications*. Wiley, New-York, 1971.
- [17] A. Friedman. *Stochastic differential equations and applications*, volume 1. Academic Press, 1975.
- [18] I. S. Gál et J. F. Koksma. Sur l'ordre de grandeur des fonctions sommables. *Indagationes Math.*, 12 :192–207, 1950.
- [19] M. I. Gordin. The central limite theorem for stationary sequences. *Soviet Math. Dokl.*, 10(5) :1174–1175, 1969.
- [20] A. Grorud et D. Talay. Approximation of Lyapunov exponents of nonlinear stochastic differential systems. 1990. Soumis pour publication.
- [21] Y. Guivarch et J. Hardy. Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov. *Ann. Inst. Henri Poincaré-Probabilités et Statistiques*, 24(1) :73–98, 1988.
- [22] G. Halász. Remarks on the remainder in Birkhoff's ergodic theorem. *Acta Math. Acad. Hungar.*, 28 :389–395, 1976.
- [23] P. Hall et C. C. Heyde. *Martingale limit theory and its application*. Academic press, 1980.
- [24] T. E. Harris. The existence of stationary measures for certain Markov processes. *Third Berkeley sump. Matht Stat. Proba.* II, 1956.
- [25] D. W. Heermann. *Computer simulation methods in theoretical physics*. Springer-Verlag, 1986.
- [26] C. C. Heyde. On the central limit theorem and iterated logarithm law for stationary processes. *Bull. Austral. Math. Soc.*, 12(1-8), 1975.
- [27] C. C. Heyde et D. J. Scott. Invariance principles for the law of iterated logarithm for martingales and process with stationary increments. *Ann. Pro.*, 1(3) :428–436, 1973.
- [28] N. Ikeda et S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North Holland, 1981.
- [29] C. T. Ionescu Tulcea et G. Marinescu. Théorie ergodique pour des classes d'opérations non complètement continues. *Ann. Math.*, 47 :140–147, 1946.
- [30] M. H. Kalos, ed. *Monte Carlo methods in quantum problems*, volume 125, *NATO ASI series*. Reidel, 1984.

- [31] Kleimann. Some exact results on stability and growth of linear parameter excited stochastic systems. M. Kohlman et V. Vogel, eds, *Stochastic Control Theory and Stochastic Differential Equations*, volume 16, New-York, 1979. Springer-Verlag.
- [32] U. Krengel. *Ergodic Theorems*. de Gruyter Studies in Mathematics, Berlin, New-York, 1985.
- [33] U. Krengel. On the speed of convergence in ergodic theorem. *Monatsh. Math.*, 86 :3–6, 1986.
- [34] P. Liardet. Regularities of distribution. *Compositio Mathematica*, pages 267–293, 1987.
- [35] J. Neveu. *Martingales à temps discret*. Masson, Paris, 1972.
- [36] J. Neveu. Potentiel markovien récurrent des chaînes de Harris. *Ann. Inst. Fourier, Grenoble*, 22(2) :85–130, 1972.
- [37] F. Norman. *Markov Process and learning Models*. Academic, New York, 1972.
- [38] S. Orey. *Lecture notes on limits theorems for Markov chain transition probabilities*. Van Nostrand, 1971.
- [39] G. Pagès. Van der Corput sequences, Kakutani transforms and one-dimensional numerical intégration. *Journal of Computational and Applied Mathematics*, 1992. A paraître.
- [40] E. Pardoux et M. Pignol. Etude de la stabilité de la solution d'une E.D.S bilinéaire à coefficients périodiques, application au mouvement d'une pale d'hélicoptère. A Bensoussan et J. L. Lions, eds, *Analysis and Optimization of systems*, volume 63, pages 92–103. Lecture Notes in Control and Info., 1984.
- [41] E. Pardoux et V. Wihstutz. Lyapunov exponents and rotation number of two-dimensional linear stochastic systems with small diffusions. *SIAM J. Appl. Math.*, 48(2) :442–457, 1988.
- [42] W. Philipp et W. Stout. Almost sure invariance principles for partial sums of weakly dependent variables. Technical report, Memoirs of the America Mathematical Society, 1975.
- [43] M. Pinsky. Instability of harmonic oscillator with small noise. *SIAM J. Appl. Math.*, 46(3) :451–463, 1986.
- [44] M. Pinsky et V. Wihstutz. Lyapunov exponents of nilpotent Ito systems. *Stochastics*, 25 :43–57, 1988.
- [45] C. Rebbi, ed. *Lattice gauge theories and Monte Carlo Simulations*, World scientific, 1983.
- [46] D. Revuz. *Markov chains*. North-Holland, 1975.
- [47] M. Rosenblatt. A central limit theorem and a mixing condition. *Proc. Nat. Acad. USA.*, 42 :412–413, 1956.

- [48] J. Rousseau-Egele. Un théorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux. *The Annals of Probability*, 11(3) :772–788, 1983.
- [49] D. J. Scott. Central limit theorems for martingales and processes with stationary increments using a Skorohod representation approach. *Adv. Appl. Prob.*, 5 :119–137, 1973.
- [50] A.N. Shiriyayev. *Probability. N° 95* , Graduate Texts in Mathematics. Springer-Verlag, New-York, 1984.
- [51] D. Talay. Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics and Stochastic Reports*, 29 :13–36, 1990.
- [52] D. Talay. Approximation of upper Lyapunov exponents of bilinear stochastic differential systems. *SIAM J. on Numerical Analysis*, 28(4) :1141–1164, 1991.
- [53] D. Talay. Approximation et simulation de solutions d'équations différentielles stochastiques. N. Bouleau et D. Talay, eds, *Probabilités numériques*, volume 10, chapitre 2, pages 23–38. INRIA, 1992.
- [54] D. Talay et L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Analysis and Applications*, 8(4) :94–120, 1990.